Psychological Bulletin

478Z 92

WAYNE DENNIS, Editor
Brooklyn College

EDWARD GIRDEN, Associate Editor (Book Reviews)
Brooklyn College

ROBERT L. THORNDIKE, Associate Editor (Statistics)
Teachers College, Columbia University

LORRAINE BOUTHILET, Managing Editor

VOLUME 50, 1953
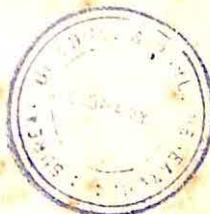
CONTENTS OF VOLUME 50

ORIGINAL CONTRIBUTIONS, GENERAL REVIEWS, DISCUSSIONS

Psychological Bulletin

VARIANCE DESIGNS IN PSYCHOLOGICAL RESEARCH

LEONARD S. KOGAN
Institute of Welfare Research, Community Service Society of New York

About a decade ago Garrett and Zubin (49), surveying applications and the potential utility of analysis of variance in psychological research design, pointed out that such techniques had not yet been widely employed. Since that time the number of psychological studies using analysis of variance has become so large that even a listing of titles would be of prohibitive length. Several statistical texts emphasizing variance analysis in psychological research have since appeared (37, 76, 101) as well as numerous methodological articles written by psychologists.[1]

The purpose of the present review is to indicate the directions and extent to which analysis of variance designs have been applied in recent psychological research. For the most part references are drawn from papers appearing in the *Journal of Experimental Psychology* and the *Journal of Comparative and Physiological Psychology* during recent years. No attempt has been made, however, to make an exhaustive survey of such applications, special emphasis being placed on papers where problem formulation, design, analysis, and inferences are presented in sufficient detail for the reader to grasp essential methodology and thus implement his understanding of experimental design

[1] Edwards (37) presents a bibliography of many of these articles

and analysis over that obtainable from the typical artificialities of statistical texts. It will be assumed that the reader is acquainted with the basic concepts and computational procedures for analysis of variance to the level of Edwards (37) and McNemar (101). Reference will be made to other readily available sources when necessary.

Problems of terminology present difficulties in any discussion of experimental designs. Psychologists have not been consistent in taking over Fisherian terminology (44, 45). While terms such as "factorial design," "latin square," "treatment," "replication," and others have gained widespread usage, terms such as "block," "plot," "varieties," etc., have apparently seemed too agronomic to be commonly used by psychologists. In the material to follow, popular terminology such as that used by Edwards (37) will be followed, with some attention being paid to alternative names which have been used.

SINGLE-CLASSIFICATION DESIGNS

Most statistical texts introduce the topic of analysis of variance by describing the partitioning of sums of squares (SS) and degrees of freedom (df) in the case of the single-classification design. The essence of this design is the presence of a single cri-

terion of classification usually represented by several independent groups of Ss upon whom the same measurements have been taken (49). The several groups typically involve the application of different experimental treatments. The usual problem to be answered by the analysis is whether the means of the several groups differ more among themselves than can be attributed to random-sampling variation from a common population. The over-all test of the significance of differences among the means is provided by an F ratio with the numerator derived from the variation of the several means and the denominator based on "pooling" the individual differences within the several groups. Other common names for this design are single-factor design, one-factor design, one-way classification, two-part analysis of variance, single-variable design, simple analysis of variance, between and within analysis, and simple classification of variates.

The most frequently used form of this design is the two-group case where the number of observations in each of the groups may be either equal or unequal. For this case the traditional method of analysis is the t test with $k(=N_1+N_2-2)df$ or the equivalent F ratio with one df in the numerator and k df in the denominator. Because of its widespread familiarity, no illustrations of the two-group case will be presented.

The extension of the single-classification design to more than two groups, despite its simplicity, is not frequently found in the literature. Franklin and Brozek (47), investigating the relationship between psychomotor performance and type of practice schedule, made use of a single-classification analysis. Thirty-six Ss were allocated to six equal groups with comparable means and standard deviations on the basis of per-

formance in "try-out" trials. The groups were then assigned different practice schedules on two psychomotor tests, e.g., three trials a day, three trials a week, etc. The single-classification design was applied in testing the over-all significance of differences among the six group means at specified trials. The analysis of variance, say, at the ninth trial appeared in the form shown in Table 1.

TABLE 1

ANALYSIS OF VARIANCE AT SPECIFIED TRIAL (47)

Source of Variation	df
Between groups	5
Within groups	30
Total	35

A slight complication appears in the single-classification design when the number of Ss in each of the several groups is unequal. The computational method of correcting for unequal N's by dividing the total squared for each group by its own N is readily found in all texts. Ammons (2) used this so-called unbalanced single-classification design in a study of rotor pursuit performance where eight unequal groups were given different conditions of pre-practice warming-up activity. As in the preceding illustration, Ammons used the design to test the over-all significance of differences among group means at specified trials. A more general example of the single-classification design with unequal N's in the groups is provided by Kelman (82) in a study involving the comparison of suggestibility scores for four groups of Ss classified as Control, Success, Failure, and Ambiguous.

Comment. The single-classification design is the prototype of the classical

experimental dictum of keeping all factors constant but the one being investigated. Reliable inference from this design demands that all conditions other than those which distinguish the several experimental groups be kept comparable from group to group or at least completely randomized among the groups. All variation over and above the differences among means is used to make the estimate of chance fluctuation or experimental error. Whenever possible, the Ss should be assigned to the several groups in a random manner. Large individual differences or heterogeneity of response among Ss within the same group enter into the estimate of error and may mask small but real differences among the groups. Failure to reject the null hypothesis, i.e., equality of the several means, is thus often attributable to small size of samples. If, on the other hand, the Ss of the experiment are kept markedly homogeneous by having them all of the same age, sex, IQ, education, etc., significant differences among groups may be found as a function of experimental variations, but the experimenter will then find it difficult to generalize from his findings to a meaningful population.

It is interesting to note that Franklin and Brozek (47) in the study cited above did not actually rely on simple randomization in selecting their six groups of Ss. Near equality of initial means and standard deviations was "forced" by distributing the Ss among the groups, not by exact pairing, but by a rough matching of high, moderate, and low scores from group to group. This attempted control of subsequent variation was not, however, taken into account in the analyses of results. It seems probable that the size of the "error variance" might have been reduced (but with the loss of 2 df) if the analysis had

been carried out for a double classification of data (see below), i.e., by adding another classification on the basis of initial score category. It should be emphasized that the writer is not questioning the conclusions of these investigators but merely using their study to illustrate the point that statistical analysis should in general conform to experimental design if maximum accuracy is to be attained.

The single-classification design is somewhat limited in efficiency because of the characteristic heterogeneity of human and animal material used in psychological research. Although this design furnishes the maximum number of *error df* for the given number of observations, the error variance is likely to be relatively large unless the several classes contain a fairly large number of observations. Frequently, a marked reduction in error variance can be gained by a slight modification of design. Perhaps the main usefulness of this design is to serve as an extension of the *t* test to more than two groups. Not only does the analysis of variance evade the practical problem of carrying out a laborious number of *t* tests when there are many experimental comparisons to be made, but it can be argued that the over-all *F* test leads to more dependable inference about possible differences among means. The basis for this argument is the increased reliability or precision of the over-all "error" term as a function of the fact that it is based on more *df* than the error based on any two subgroups. Moreover, such *t* tests are not independent and "significant" *t*'s tend to be found more frequently than indicated by the chosen level of confidence. Thus even when all samples have actually been chosen at random from the same population, separate *t* tests often indicate apparent significance of differences. With

six samples, for example, Cochran and Cox (27, p. 18) state that the observed t between the highest and lowest mean will exceed the tabled .05 level about 40 per cent of the time.

Multiple-Classification and Factorial Designs

In the single-classification design it is possible to increase the "sensitivity" of the experiment, i.e., allow the detection of smaller differences among the experimental groups, by using a greater number of cases or by improving the reliability of measuring the dependent variable under consideration. A third technique for increasing the sensitivity of an experiment is by deliberately arranging the design so that known sources of variability can be controlled and separated both from the experimental comparisons and from the estimate of experimental error. One of the major purposes of multiple classification in modern experimental design is to provide methods for minimizing experimental error by the control and isolation of extraneous sources of variation. Perhaps the simplest example of such a controlled arrangement is the method of pairing cases. The reduction of the standard error of difference between the means of paired samples, when the pairing results in significant positive correlation between the samples, illustrates the basic procedure of increasing the sensitivity of an experiment by multiple classification. Here the use of the correlation term in the standard error formula or the equivalent method of analyzing the distribution of differences between paired scores is exactly the same as breaking down the total variation of scores into the three mean squares: between treatments, between pairs, and residual.

The general principle involved in the pairing of cases is to increase the homogeneity of experimental material by employing the arrangement known in experimental agriculture as *randomized blocks* (45). This design, as the name implies, consisted originally of the marking out of blocks of land with each experimental treatment then being randomly assigned to plots within each block. Each block is often referred to as a replicate. The resulting yields can then be entered in a two-way table with rows representing the treatments and columns representing the blocks. Analysis of variance separates three sources of variation: treatments, blocks, and error. The psychological analogue to the randomized block is seen to be either the single S who receives all experimental treatments in randomized order or a group of comparable Ss, each of whom is randomly assigned to one of the experimental variations. In animal experiments the block may consist of litter mates, thus allowing the control of variation due to strain, age, weight, etc., while in experiments with humans it is common to form the block on the basis of sex, IQ, socioeconomic level, initial scores on the dependent variable, etc. Many examples of the use of randomized blocks in multiple-classification design will be presented below. In all cases a priori information is used in an attempt to increase the precision of experimental comparisons by removing extraneous sources of variation.

Another basis for multiple classification in experimental design is represented in the so-called factorial design. In this case the investigator is interested in studying the effects of a number of different experimental factors, each of which is varied in two or more ways. The experimental treatments of the factorial design involve all possible combinations of the factors under consideration. In dis-

tinction to the classical rule of holding all but one factor constant, the factorial experiment depends on the simultaneous variation of as many factors or conditions as the experimenter chooses to control. Not only is it usually difficult to keep other relevant conditions constant as demanded by the classical single-factor design, but even if such control were attained the basis of generalization would accordingly be limited to the particular pattern of constancies maintained in the given study. Fisher (45) stresses the greater efficiency and comprehensiveness of the factorial study. Efficiency is derived from the fact that several factors may be evaluated with the same precision and by fewer observations than would be the case in carrying out separate studies for each factor. Greater comprehensiveness comes from the possibility of evaluating not only the over-all effects of each of the factors but their interactions as well. A broader basis of inductive generalization is derived from the consideration that each factor is evaluated, not with other factors kept arbitrarily constant, but over the range of variation of the other factors involved in the experiment. Because of these unique properties, psychological experimentation is becoming increasingly characterized by the use of factorial designs, often in combination with the principle of randomized blocks.

The distinction between multiple-classification designs and factorial designs in psychological research (32, 34, 49) is sometimes difficult to make. Baxter (6) has presented a discussion of this distinction. If a given rubric of classification can be taken to represent variation either on a quantitative scale, i.e., different amounts, degrees, or levels of a variable, or on a qualitative continuum, i.e., differ-

ent categories of a set of experimental conditions or treatments, the particular classification may be called a "factor." On the other hand, if the axis of classification does not represent a quantitative or qualitative variable, e.g., subjects, months, schools, that particular classification would not usually be referred to as a factor in strict parlance. It should be emphasized, however, that this literal conception is not widely adhered to and the term "factorial design" is rather loosely employed, not only by psychologists but also by many statisticians. In any case the analysis of multiple-classification and factorial designs generally involves analogous procedures.

The designs falling in the category of multiple classification are most simply referred to in terms of the number of classifications of the data or in terms of the number of factors involved. Thus one may refer to two-way, three-way, etc. classifications or two-factor, three-factor, etc. designs. Other terms which are sometimes used are complex design, three-part, four-part, etc. analysis of variance or, simply, higher-order classifications.

There are several major subcases in multiple-classification and factorial design. The simplest case is that in which there is but one replication, each subclass containing a single observation. In this case the estimate of experimental error is provided by the highest order interaction term. The second case is the design where the subclasses of the multiple classification or each unique factorial combination contain equal numbers of observations. The third case entails frequencies in the subclasses which are proportionate with the marginal totals. And, finally, there is the complex case where the subclasses contain unequal and disproportionate

numbers of observations. Examples will be provided below for each of these variations in fundamental design.

Double classification with one observation per subclass. Carpenter (17) carried out a study of the effect of prolonged visual search, submitting his results as evidence that rate of blinking can be used as a criterion of visual efficiency. Twenty Ss were engaged in a visual task (Mackworth's Clock Test) where they responded to a specified cue by pressing a key twelve times during each half-hour. The measure analyzed was the number of eyeblinks during a two-hour run. The mean number of blinks per minute in each half-hour was calcu-

lated for each S and these means were treated as single observations. The analysis appeared as in Table 2.

It should be noted in Table 2 that the error estimate is actually based on the interaction between half-hours and Ss. This example illustrates the general form of the double-classification design where there is one observation in each subclass, but in this case the Ss cannot be regarded as "randomized blocks" since the columns represent successive periods of time and not a random arrangement of different experimental conditions. Although such a refinement was not apparently necessary in this study to demonstrate the "significant" increase in blinking rate, in some cases it may serve to make the comparison of successive time periods more sensitive if individual variations in time regression are taken out of the "error" term as described in the section on "Repeated Measurements" presented below.

Similar two-way classifications, one axis representing Ss and the other based on successive periods of time, were used by Siegel and Stuckey (122) in a study of the diurnal course of water and food intake in rats. The use of a double-classification design with Ss operating as randomized blocks is found in a study by Chapanis, Rouse, and Schachter (18) of the effects of various kinds of intersensory stimulation on form discrimination at low brightness. Although only three Ss were used in this latter study, with each S receiving a random arrangement of six experimental conditions, the design illustrates how an overwhelming amount of consistent individual differences may be separated from the estimate of experimental error by the use of Ss as randomized blocks. Another example of a double-classification design with Ss as one criterion of classification is

TABLE 2

MEAN NUMBER OF BLINKS PER MINUTE
FOR EACH S IN EACH
HALF-HOUR (17)

Sub-ject	Blink Rate per Minute				
	Half-Hours				Mean
	1	2	3	4	
1	8.4	16.9	16.2	17.2	14.7
2	7.3	14.2	15.4	16.6	13.4
3	10.0	15.6	19.4	19.6	16.2
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
18	12.2	16.9	18.2	19.7	16.8
19	59.1	60.5	42.2	83.7	61.4
20	11.1	21.9	16.9	30.3	20.1
Mean	17.4	21.9	21.1	24.8	21.3

Analysis of Variance

Source of Variation	df
Between half-hours	3
Between subjects	19
Residual (error)	57
Total	79

provided by Postman (111) in an experiment relating the efficiency of recognition of nonsense syllables to number of correct and incorrect items in the recognition tests. Double-classification designs with three Ss as one axis of classification were also used by Mann and Passey (100) in a study of adjustment to the postural vertical as a function of magnitude of tilt and duration of exposure. Although this study was factorial in design (8 durations of exposure time and 6 variations of tilt), the investigators neglected the opportunity of evaluating possible interaction between tilt and exposure time by treating the two factors in separate double-classification analyses.

Double classification with equal numbers of observations per subclass. In this design the double classification is replicated so that there are equal numbers of observations within the subcells. This availability of replication allows the "interaction" term used as "error" in the preceding design to be itself tested against the residual "within cells" mean square. In some studies the experimenter may be particularly interested in possible interaction effects and it is impossible to make a judgment about the possible significance of such effects without some form of replication. Chapanis and Leyzorek (19) employed this design in a study on accuracy of visual interpolation. Eleven Ss were given randomly arranged trials where the task was to estimate the position of stimuli by means of 11 different numerical scales. Standard deviation scores based on 25 estimates with each of two different instruments were computed for each S for each scale and the two resulting scores were treated as replications within the subclasses.[2] The form of analysis is indicated in Table 3. Although the investigators

chose to consider the two scores within each subclass as simple replications leading to the analysis presented in Table 3, a somewhat more informative analysis might have been made by treating the experiment as a triple-classification design with Ss as one axis of classification. Since each S was tested "randomly" on the same two instruments, it would appear that the instruments could be used as a third axis of classification in the form presented below in Table 4. If this had been done, the total df would have been allocated as follows: 10 df for Ss, 10 df for scales, 1 df for instruments, 100 df for interaction between Ss and scales, 10 df for interaction between Ss and instruments, 10 df for interaction between scales and instruments, and 100 df for the triple interaction.

Two-factor designs with equal numbers of observations in the subcells. This design involves the investigation of the effects of two factors, each of which is varied over a designated number of levels. Equal numbers of different Ss are randomly assigned to each of the several factorial combinations. Whereas in the preceding double-classification design with Ss as one of the axes of classification the differences between the effects of the experimental treatments were associated with intrasubject variation, in this design differences in treatment effects are associated with intersubject variation. The basic estimate of experimental error is derived from differences in response of Ss subjected to the same experimental conditions. The presence of possible interactions between the two factors

---

[2] It should be noted that this example consists of an analysis of variance of a set of sample standard deviations. Bartlett (5) recommends that the analysis be carried out with a logarithmic transformation of variances in such cases.

TABLE 3

STANDARD DEVIATIONS OF RELATIVE ERRORS OF ESTIMATION
FOR EACH SUBJECT AND EACH CONDITION (19)

(The entry in each cell is based on 25 estimates.)

Subject	Instrument	Number Scale					Mean
		1000	2000	· · ·	10000	2.5	
A	1	3.96	2.74	· · ·	3.42	2.74	4.49
	2	3.95	3.36	· · ·	3.19	4.48	
B	1	3.36	2.67	· · ·	3.34	6.62	
	2	4.72	3.86	· · ·	3.66	10.64	4.74
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
K	1	2.73	5.58	· · ·	3.15	2.89	
	2	3.85	3.57	· · ·	2.59	3.10	4.28
Mean*		3.99	4.63	· · ·	3.89	7.32	5.14

Analysis of Variance

Source of Variation	df
Between subject means	10
Between scale means	10
Subject-scale interaction	100
Between instruments within cells	121
Total	241

* Mean, Instrument 1 = 5.21; Mean, Instrument 2 = 5.08.

may be evaluated when this design is used. Analytic procedure is the same as for the double-classification design with equal numbers of observations in the subcells. Many examples of this so-called replicated two-factor design were found in the literature.

Kimble and Bilodeau (85) employed a 2×2 factorial design in a motor learning study in which initial and final scores on the Minnesota Rate of Manipulation Test were analyzed as a function of two conditions of work and two conditions of rest, with 24 Ss in each of the four possible combinations of conditions.

Other examples were a 2×3 design with eight Ss per combination used by Norris and Grant (108) in a study of eyelid conditioning as a function of inhibitory or passive instructions and three conditions of reinforcement; a 2×2 design with six Ss per cell used by Lawrence and Miller (89) in investigating resistance to extinction as a function of two variations in number of reinforced trials and two amounts of reinforcement; a 4×4 design with five Ss per cell applied by Grant and Schneider (60) in a study of the magnitude of GSR response during extinction as a function of

four levels of CS intensity during both reinforcement and extinction; a 4×4 design with four Ss per cell used by Grant and Schneider (59) in studying the relation of intensity and frequency of a conditioned eyelid response during extinction to four variations in intensity of CS during reinforcement and extinction; a 3×3 design with ten Ss per cell by Chernikoff and Brogden (20) in a study of the effects upon sensory conditioning of three variations in pretraining treatment and three types of instructions; and a 2×2 design with 20 Ss per cell used by Grant, Norris, and Boissard (57) in studying the change in mean magnitude of eyelid response from pretest to posttest as a function of the presence or absence of dark adaptation and the presence or absence of pseudo-conditioning reinforcement.

Double-classification or two-factor designs with unequal but proportionate numbers of observations in the subclasses. This design differs from the usual two-way classification described above in that the numbers of observations in the subclasses, although not the same, are proportionate for each row and column to the numbers of observations in the marginal totals. For example, a 2×2 table with one row having subclasses containing two and four observations and the second row containing three and six observations would fit this description. The analysis of variance for this design offers no computational difficulties since the SS for rows, columns, row by column interaction, and within subclasses are additive to the total SS. The only corrections necessary for the unequal entries in the subclasses are the same as those used in the analysis of the single-classification design with unequal numbers of cases in the several groups (124). Webb (130) used this

design in a study of the strength of a food-reinforced response as a function of varying conditions of an irrelevant drive. The irrelevant drive in this case was thirst, and one classification of his data was based on the setting up of four independent groups with different periods of thirst deprivation. Each of these groups contained a total of 18 rats. The second classification was based on differentiating the sex of the S. The unequal but proportionate subclass frequencies resulted from the fact that each group consisted of 10 males and 8 females.

In his analysis Webb made the assumption that the within-subclass mean square provided the appropriate estimate of error variance for testing the over-all significance of differences among the four major experimental groups. This was a warranted procedure since in all of the measures analyzed (latency, extinction) the F ratios of group-by-sex-interaction mean square to within-subclasses mean square were "nonsignificant." In some studies, however, the experimenter might find that the "interaction" is "significant," and he may then desire to test the intrinsic effect of the main classification under the assumption that the appropriate error term should include a compounding of both interaction variation and within-subclass variation. In the ordinary case where the numbers of entries in each subclass are equal, this is done simply by forming an F ratio of main effect mean square to interaction mean square. In the present design, however, where the numbers of observations in the subclasses are unequal but proportionate, Smith (123) has recently called attention to a qualification in procedure when the investigator desires to test the significance of a main effect over and above the

variation due to possible interaction. The appropriate method for carrying out this test of significance is somewhat complex and involves setting up an F ratio consisting of multiple terms in both numerator and denominator.

An example of a two-factor design with proportionate subclass N's is also reported by Kelman (82). As in Webb's study, the interaction was not found to be significant and no complications developed in testing the main effects.

Double-classification or two-factor designs with disproportionate numbers of observations in the subclasses. The situation sometimes arises in experimentation or investigation where the numbers of observations in each of the subclasses of a multiple-classification design are not only unequal but also disproportionate with the marginal totals. In such cases the simple corrections for unequal subclass frequencies which are applied in single-classification designs or multiple-classification designs with proportionate frequencies are no longer adequate. Such a state of affairs may arise because of various reasons, e.g., failure of .Ss to meet appointments, loss of animals, type of investigation, etc. Such designs are referred to as the "nonorthogonal case" because the estimates of variance computed for the several sources of variation are interdependent (124). Thus in a 2×2 classification if one were to calculate separately the SS for columns, rows, column-by-row interaction, and residual he would find that these SS would not generally add to the total SS.

In the simplest case only one or two items of data may be missing from some of the cells. The common method for estimating a small number of missing entries and filling out a table was developed by Yates (132) and is readily accessible in Snedecor (124), Anderson(3), and Cochran and Cox (27). The general problem of analyzing tables of multiple classification with disproportionate subclass numbers is discussed by Lindquist (95) and Johnson (76), but in the absence of specific cautions, students referring to McNemar (101) and Edwards (37) may incorrectly infer that the corrections described for single-classification inequality of frequencies are sufficient. A number of different solutions to the problems of disproportionate frequencies have been proposed, all of which involve approximations based on varying assumptions. Snedecor (124) has presented a comprehensive summary of the so-called methods of fitting constants, unweighted means, expected subclass numbers, and weighted squares of means. In these methods it is generally assumed that the usual within-subclasses SS furnish an appropriate estimate of error variance. Tsao (128), on the other hand, has derived solutions where this assumption is not made. Two of the basic decisions which the investigator must always make in selecting a solution are whether or not interaction is "significant" and whether or not disproportionality is characteristic of the inferred population.

In employing multiple-classification designs with disproportionate subclass frequencies, some psychologists have taken cognizance of the special methods necessary for this case while other studies have been reported in which no apparent adjustments were made. Bray (11) used corrections suggested by Snedecor (124) in analyzing conformity scores in an autokinetic situation of 2×2 design, where unequal numbers of Ss were classified according to racial attitude and whether or not the confederate was a member of a specified

race. Porter, Stone, and Eriksen (110) also used Snedecor's methods in analyzing 2×3 and 2×9 designs in a study where maze error scores were being compared for rats given electroconvulsive shocks in late infancy and control litter mates. Some studies in which the analysis apparently failed to take adequate account of disproportionate subclass frequencies were a 2×2 design by Jenkins and Postman (75), a 3×3 design by Postman and Jenkins (112), a 2×10 design by Hunt, Schlosberg, Solomon, and Stellar (72), and 2×2 and $3 \times 2 \times 2$ designs by Citron, Chein, and Harding (24).

In some experiments the investigator has sought to evade the problem of disproportionate subclass frequencies by ignoring the individual observations in the cells and analyzing the data as if there were no replication, i.e., analyzing subclass means as if they were single observations. In general this procedure cannot be rigorously defended, especially when the frequencies are markedly dissimilar, since such means are differentially reliable and nonorthogonality is still inherent in the data.

Triple-classification designs with subjects as a criterion of classification. In many essentially two-factor designs a third axis of classification is provided by the fact that each S undergoes all of the experimental variations or conditions, frequently in random order. Because "between subjects" is considered a major source of variation in such designs, there is no "within subclasses" estimate of error and the basic estimate of experimental error is provided by the triple- or second-order interaction mean square. Such a design was used by Solomon (125) in a study of the effect of effort upon distance discrimination, where ten rats went through four successive experimental

sessions, alternately running a maze with and without a load over a period of eight days. Analysis of the ordinal number of the side alley first entered during each session followed the form of Table 4. In this case the Ss cannot be regarded as "randomized blocks" since each received the same sequence of experimental variations.

By way of didactic comment about Solomon's analysis, the 9 *df* for rats might have been separated into one *df* for sex and 8 *df* for rats within sex groups. Possible sex difference might then have been evaluated by means of an *F* ratio derived from these two sources of variation. Furthermore, it should be noted that days (not analyzed) are confounded with sessions. Finally, the comparison of effort levels is also confounded with days since performance under the condition of "load" as a whole took place one day later than performance without the load. In this case, however, the general temporal trend was to enter a more remote alley and this was an opposing trend to the tendency exhibited under load. One would thus predict that the apparent difference between effort levels might have been even greater, had the two experimental conditions been randomized for each S.

Littman (97) used a similar $4 \times 4 \times 11$ design with two groups of 11 Ss in a study of the generalization of a conditioned GSR to tones other than the original CS. Other applications of this design were made by Black (10) in a $5 \times 2 \times 25$ study of intensity of oral responses to two types of messages under five levels of intensity, and by Beebe-Center, Black, Hoffman, and Wade (7) in a $3 \times 12 \times 9$ investigation of per diem consumption as a measure of preference in the rat.

Trifactorial designs with one observation per subclass. The analysis of the triple-factor design with one ob-

TABLE 4

ORDINAL NUMBER OF SIDE ALLEY FIRST ENTERED BY RATS DURING THE EIGHT TEST SESSIONS (125)

	Session—Pairs									
	1		2		3		4			
	Session (day)									
	70	71	72	73	74	75	76	77		
	Level of Effort								Mean	
Rats	Normal	Load	Normal	Load	Normal	Load	Normal	Load	Normal	Load
♂ 1	3	3	3	2	4	4	4	3	3.5	3.0
2	4	3	3	3	5	5	4	5	4.0	4.0
3	4	3	4	2	5	4	4	4	4.3	3.3
4	3	1	3	2	4	2	4	2	3.5	1.8
5	3	3	4	4	4	4	4	3	3.8	3.5
♀ 1	4	1	4	3	4	3	3	3	3.8	2.5
2	2	1	2	1	1	1	2	1	1.8	1.0
3	3	3	4	3	4	3	4	4	3.8	3.3
4	2	2	4	2	4	3	3	1	3.3	2.0
5	3	2	5	4	3	2	4	3	3.8	2.8
Day Means	3.1	2.2	3.6	2.6	3.8	3.1	3.6	2.9	3.5	2.7
Session-Pairs	2.7		3.1		3.5		3.3			

Analysis of Variance

Source of Variation	df
Rats (R)	9
Effort levels (E)	1
Sessions (S)	3
R×E	9
R×S	27
E×S	3
Error (R×E×S)	27
Total	79

servation per subclass is analogous to that shown in Table 4, with the replacement of rats, i.e., subjects, by the third factor. Helson (70) utilized this design in analyzing a 2×4×10 factorial experiment where time errors with handwheels were classified according to wheel diameter, amount of friction, and speed of turning. Actually, in this study subcell values were averages for different groups of six Ss each.

Trifactorial designs with replications. The presence of equal numbers of observations in the subclasses of a three-factor design affords a within-cells residual which can be used in testing the significance of the triple interaction. Wilson (131) used this design in a study of the frequency of remote associations at recall for rote learning. His application, using three *S*s in each combination of a 4×4×3 design, yielded an analysis as in Table 5.

Gebhard (50) used a similar design in a 2×2×2 study investigating attractiveness rankings of tasks classified according to experience (success-failure), expectation of task difficulty, and strength of need. Other investigators using this design were Grant (55) in a 2×2×2 factorial study of responses to a card sorting task; Grant, Hornseth, and Hake (65) in a 2×2×2 study of the influence of intertrial interval on the Humphreys' effect with verbal responses; and Grant and Mote (63) in another 2×2×2 study of the effects of brief flashes of light upon dark adaptation. Lawrence (88) reported a study involving a 2×2×2 design in which certain comparisons were confounded because of the nature of the experimental design. A study by Conklin (28) which apparently involved a three-factor design in an investigation of the effects of temperature, duration of session, and adaptation on skin resistance presents an allocation of *df* which is difficult to reconstruct.

The problem of disproportionate subclass frequencies with tables of multiple classification again arises in this design. In a 2×2×5 design used by Bendig and Braun (8) for studying maze behavior, adjustments were made both for missing cell entries and for differing subgroup *N*'s according to suggestions by Snedecor (124),

Anderson (3), and Schoenfeld (121). However, in a 2×4×2 study by Newman and Scheffler (107) concerned with sex differences in emotional reaction to the news, where sex, educational level, and type of

TABLE 5

FREQUENCY OF REMOTE ASSOCIATIONS AT RECALL (131)

Interval between Learning & Recall (mins.)	Spacing between Trials (secs.)	Degree of Learning (% of perfect anticipation)			
		50	75	100	200
0	6	—*	—	—	—
	30	—	—	—	—
	60	—	—	—	—
2	6	—	—	—	—
	30	—	—	—	—
	60	—	—	—	—
5	6	—	—	—	—
	30	—	—	—	—
	60	—	—	—	—
20	6	—	—	—	—
	30	—	—	—	—
	60	—	—	—	—

Analysis of Variance

Source of Variation	df
Degree of learning (D)	3
Intervals following learning (I)	3
Conditions of spacing (C)	2
D×I	9
D×C	6
I×C	6
D×I×C	18
Within cells	96
Total	143

* 3 *S*s per cell; data not provided.

newspaper were treated as major sources of variation, there is no evidence that account was taken of the markedly disproportionate frequencies in the subclasses.

Quadruple and higher classification designs. These designs represent

further elaboration of the principles already described. In some studies all of the classifications can be regarded as factors while in other studies one of the classifications of the data depends upon the fact that each S undergoes every variation of experimental combinations. Quadruple-classification $4 \times 2 \times 2 \times 2$ designs were used by Preston, Spiers, and Trasoff (116) in a level of aspiration study. Grant, Hornseth, and Hake (61) applied a $5 \times 2 \times 4 \times 40$ design, with Ss as one criterion of classification, in a study of sensitization of the beta-response to visual stimuli. Littman (98) used a $3 \times 2 \times 2 \times 6$ design in a latent learning experiment, while Horowitz (71) applied several $10 \times 4 \times 4 \times 2$ designs with Ss as one classification in a study of visual acuity. Child and Grosslight (23) made use of a $3 \times 2 \times 2 \times 2$ factorial design in a study of substitute activity with the added complication of breaking down one of the factors into a major and minor subclassification. A five-way $7 \times 10 \times 2 \times 4 \times 14$ classification was used by Kuntz and Sleight (87) in a study of legibility of numerals as a function of height/width ratio, type of numeral, background, and brightness. The highest number of criteria of classification found in the literature surveyed was applied by Licklider, Bindra, and Pollack (94) in a study comparing the intelligibility of normal and "square" speech. Two "talkers" and two "listeners" furnished two of the major criteria of classification in a $2 \times 5 \times 2 \times 2 \times 3 \times 10$ design. The authors present an interesting argument for the rationale of generalizing from such a small number of Ss.

Comment. This section has dealt with the possibilities for increasing the precision and scope of experiments by use of randomized blocks and factorial design. In planning an experiment involving the comparison of the effects of several experimental variations, the investigator must always decide whether to use the same, matched, or different Ss for the various treatments or treatment combinations. If the same or matched Ss undergo all treatments in randomized order as in the usual factorial design, it is often possible to increase the precision of experimental comparisons by removing variation associated with over-all differences among such "blocks." Assuming that the total number of observations is the same, such an advantage must be weighed against the broader basis for generalization which is derived from the use of a larger number of randomly assigned Ss. In experiments where naïveté is essential for Ss undergoing a given treatment it is obvious that the design should contain different Ss in each of the subclasses. Similarly, wide individual variations in practice or fatigue effects in the design where each S undergoes all experimental combinations would tend to result in marked interactions between Ss and treatments, thus tending to obscure differences in the main effects of the several factors. If temporal variation is itself a main subject of investigation little would be gained from the conclusion that Ss show consistent temporal trends when each S has undergone several experimental treatments in randomized order. The following section on "Repeated Measurements" will present some common useful designs when temporal trend is a main topic of study.

Factorial designs, often involving a fairly sizable number of factors, have become very prominent in recent psychological research.[3] In the main,

---

[3] Edwards and Horst (40) have facilitated the computations involved in higher-order multiple-classification designs by furnishing a method for the direct calculation of second-order and higher interaction SS.

such designs have been a boon to experimental methods because they allow the systematic, economical exploration of the effects of a number of different factors as well as possible interactions among the factors. Programs of research, sequentially investigating the effects of varying one experimental factor at a time, such as characterized the field of learning in the past, can be immeasurably hastened and increased in generality by the application of factorial designs. On the other hand, there seems to be a tendency on the part of some experimenters to sacrifice considerations of sample size, representativeness of samples, and both reliability and validity of measurement in their enthusiastic endeavor to test large numbers of hypotheses by means of factorially designed experiments. At the extreme, for example, a complex factorial study providing many *df*'s for making many tests of significance might be carried out for a single *S*. Multiple observations could be secured for each subclass by measuring the dependent variable several times for each treatment combination. The precision of such an experiment might be very high and conclusions valid for the unique *S*, but who would attempt to generalize from the results, whether null hypotheses were rejected or not? The writer discovered no instance of the use of a single *S* in factorial design, but many investigators have reported experiments in which broad inferences were drawn from less than a half dozen *S*s.

Although, in principle, there is no limitation on the number of experimental factors which may be involved, difficulties frequently arise in the interpretation of complex factorial designs. The number of treatment combinations increases very rapidly and often limitations in apparatus or other circumstances cause a large-scale experiment to stretch

out over a considerable period of time. The classic example of a fairly elaborate factorial experiment in psychological research is that of Crutchfield (**32**, **33**). In this study the topic of investigation was string-pulling in rats as a function of five factors, each varied over three levels. A single animal was assigned to each of the 243 treatment combinations. Complete analysis would yield a list of 31 mean squares: 5 main effects, 10 two-factor interactions, 10 three-factor interactions, 5 four-factor interactions, and 1 five-factor interaction. In such a case it is generally assumed that interactions involving three or more factors can be "pooled" to provide an adequate estimate of experimental error. Fisher (**44**) describes this procedure of dispensing with absolute replication in estimating error as the method of "hidden replication" and points out the possibilities of loss of precision in tests of significance when high-order interactions are not really negligible.

Whether high-order interactions can ordinarily be assumed to be unimportant in psychological research is problematic, but it is certain that they cannot be evaluated when there is no replication. Furthermore, even with replication they frequently present puzzling problems of interpretation to the experimenter and, since large psychological studies are rarely repeated, there is little opportunity to compare their consistency over a series of experiments.

Psychologists in general have not paid much attention to the practical and experimental advantages of "confounding" in the planning of experiments. Confounding in this connection refers to the deliberate arrangement of the experiment so that certain mean squares represent the effects of more than one known source of variation. Experimenters often go to great length to avoid the

possibility of confounding experimental factors, sometimes to the considerable enlargement of their studies, even when previous studies have fairly well demonstrated that the factors concerned do not interact. Of similar character is the practice of running all possible combinations in a factorial experiment and then combining high-order interactions to estimate experimental error. The basic principle of deliberate confounding is to use "incomplete blocks," i.e., blocks within which all treatment combinations do not occur (45). In general, the purpose of such confounding is to increase the precision of selected experimental comparisons while sacrificing the possibility of evaluating other comparisons, e.g., high-order interactions.

A simple illustration will clarify the basic idea in deliberate confounding. Let us suppose that we have a three-factor experiment, each factor being varied over two levels. Representing factors by letters and levels by subscripts the eight possible combinations may be separated into two subgroups: (a) $A_1B_1C_1$; $A_1B_2C_2$; $A_2B_1C_2$; $A_2B_2C_1$ and (b) $A_1B_1C_2$; $A_1B_2C_1$; $A_2B_1C_1$; $A_2B_2C_2$. In the usual complete factorial experiment where each S serves as a block, every S would undergo all eight experimental treatments. Let us, however, modify the design so that five Ss undergo all the combinations listed after (a) while five other Ss undergo those listed after (b). Each S would then represent an incomplete block. The resulting analysis of variance would then allot 1 df each to A, B, C, AB, AC, and BC; 9 df to Ss; and 24 df to the error estimate.

The single-factor and two-factor effects are not influenced by differences among Ss (blocks) while the three-factor interaction ABC is completely confounded with these differ-

ences. In the usual "complete" experiment this latter interaction would have been estimated from the difference of (a) and (b) above. In such a case the decision to employ the confounded design might be based upon the fact that each S is available for only half of the experimental sessions, a desire to avoid fatigue or boredom on the part of Ss, or any other reason which might justify halving the experimental period for each S. The aim in this particular design is to reduce the error variance used to test the significance of the main effects and two-factor interactions by sacrificing the second-order interaction.

The reader should not conclude that confounding is possible only when individual Ss serve as blocks. Confounded designs may at times be fruitfully employed when each S undergoes only one experimental combination. Nor are such designs limited to the confounding of high-order interactions. Baxter (6) has discussed various possibilities for increased precision in experimental research through the use of confounding. In the main, however, the most comprehensive presentations of designs involving deliberate confounding are found in sources dealing primarily with experimental agriculture (27, 45, 83, 124). The following are some hypothetical examples of situations where the experimenter might consider the possible advantages of confounding by means of "incomplete" blocks:

1. The Ss fall into homogeneous groups, e.g., by sex, age, family, IQ, school, visual acuity, etc., but there are insufficient Ss in each group to allow carrying out all treatment combinations. A common example is the limited size of groups of litter mates.

2. Several experimenters might simultaneously handle portions of the entire program, thus speeding up completion of

the experiment. General differences in response of Ss due to the experimenters could be removed by confounding the blocks (experimenters) with unimportant interactions. The same principle might be applied when different machines are used to present experimental stimuli or in experiments which involve the use of confederates.

3. In some experiments time and space considerations might determine the separation of blocks. Several different experimental rooms may be involved or relevant environmental conditions may vary from day to day. Treatment combinations belonging to a block could be compared with greater precision than would be possible if such sources of heterogeneity were ignored.

In all of these cases it is presumed that there is significant block-to-block variation with respect to the dependent variable being studied.

REPEATED-MEASUREMENTS DESIGNS

In many studies the experimenter is especially interested in analyzing successive changes in measures obtained repetitively from one or more groups of Ss. Some of the types of investigation in which the problem of repeated measurements arises are (*a*) examination of learning and extinction data; (*b*) studies of dark adaptation; (*c*) investigations of performance and fatigue; (*d*) analysis of sequential measures of physiological or sensory-motor functions for varying treatment groups. The repeated-measurement situation is so common in psychological research that Edwards (37) devotes an entire chapter to this topic. Several articles have been largely devoted to this type of design (1, 86, 96).

Single group with repeated measurements. If there is but a single group of Ss, the investigator may be primarily interested in determining whether the group in general shows a

significant trend during the successive trials or periods. The simplest method of analysis is a double-classification design where rows represent different Ss and columns represent successive trials. If the *F* test for trials mean square over the Ss × trials interaction mean square is "signifi-

TABLE 6

MEAN MAGNITUDE (Mm) OF CRs AVERAGED FOR SUCCESSIVE FIVE-TRIAL BLOCKS DURING FIRST DAY REINFORCEMENT TRIALS*
(108)

Sub-ject	Successive Five-Trial Blocks				
	1–5	6–10	· · ·	36–40	41–45
1	—	—	—	—	—
2	—	—	—	—	—
3	—	—	—	—	—
4	—	—	—	—	—
5	—	—	—	—	—
6	—	—	—	—	—
7	—	—	—	—	—
8	—	—	—	—	—
Mean	0.00	0.00	· · ·	.67	.37

Analysis of Variance

Source of Variation	df
Group slope	1
Between individual means	7
Between individual slopes	7
Individual deviations from linearity	56
Total	71

* Data in body of table not provided.

cant," one concludes that there is nonrandom trial-by-trial variation, i.e., trial means are not the same (see Table 2). Such an analysis, however, does not indicate whether the trial means follow a regular linear or curvilinear trend. In order to "test" for the presence of a consistent trend, one must fit curves to the data, taking into account both individual and

group regressions upon the time scale. One of the methods suggested by Alexander (1) was applied by Norris and Grant (108) in a study of eyelid conditioning to test the statistical significance of group slope in a design involving nine successive five-trial blocks for a group of eight *S*s. The measure analyzed was mean magnitude of a CR and the analysis appeared as in Table 6. Alexander (1) points out that in some cases apparent "significance" of group slope may be attributable to wide variations in individual slopes.

Independent groups with repeated measurements. A more complex case with respect to repeated measurements occurs when several independent "treatment" or "methods" groups are involved and the investigator wishes to compare the trends exhibited by the several groups. If the assumption, among others, is made that individual regressions are parallel, analysis is readily made in terms of a double-classification design with the between-*S*s variation being subclassified into a between-treatments source of variation and a between-*S*s within-treatment-groups source of variation (37, 86, 96). This procedure of decomposing composite classifications or variables "nestled" within other variables will frequently prove to be valuable in the complete analysis of many experimental designs (cf. the "split plot" design described in detail by Cochran and Cox [27]). This form of analysis was used by Liberman (93) in analyzing 8-trial acquisition and extinction trends for two groups of 24 rats in a study of transfer effects. The variance breakdown is shown in Table 7.

Similar analyses of repeated measurements for independent groups were carried out by Furchtgott (48) in a study of maze swimming for three groups of rats exposed to different levels of X-irradiation, and by Bernberg (9) in comparing the effects of shock and narcosis upon maze-learning ability in young rats.

The analysis of repeated measurements for independent groups takes on a more complex form if possible individual and group variations in linear regression are taken into account. Alexander (1) provides an

TABLE 7

LOG LATENCIES FOR ACQUISITION OF THE
RUNNING RESPONSE* (93)

Rats	Acquisition Trials							
	1	2	3	4	5	6	7	8

Group A: Running First, Bar-Pressing Second

1	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	—
.								
.								
.								
23	—	—	—	—	—	—	—	—
24	—	—	—	—	—	—	—	—

Group B: Bar-Pressing First, Running Second

25	—	—	—	—	—	—	—	—
26	—	—	—	—	—	—	—	—
.								
.								
.								
47	—	—	—	—	—	—	—	—
48	—	—	—	—	—	—	—	—

Analysis of Variance

Source of Variation	df
Trials (T)	7
Between groups (G)	1
Between *S*s in same group (S)	46
Interaction: T×G	7
Interaction: T×S	322
Total	383

* Data not provided.

analytic procedure for such a trend analysis. His suggestions were applied by Grant, Riopelle, and Hake (64) in comparing extinction trends for three groups of 15 *S*s, each group being given a different reinforcement pattern for an eyelid CR. A trend analysis of CR magnitude scores was carried out for five successive blocks of trials with the analysis taking the form shown in Table 8.

Similar analyses were found in other studies by Grant, Hake, and Schneider (58) and Grant and Norris (56).

Repeated measurements in multiple-classification designs. Many variations of the repeated-measurements design may occur. The within-subclasses mean square of a factorially designed experiment may be based on successive measurements of the same *S*s, or several *S*s within the same subclass may be repetitively measured. In a sense, all designs where the same *S*s undergo several experimental variations are applications of the "repeated-measurements" principle, although the term has generally been used to refer to the case where the effect of the same treatment is measured successively over a period of time. Whereas in the usual factorial experiment with *S*s as one axis of classification the *S*s undergo the several experimental combinations in randomized order, it may sometimes be necessary to have all *S*s undergo the same sequence of treatments. In other experiments the *S*s may undergo the treatments in differing orders, but the investigator may wish to eliminate a general temporal effect, e.g., transfer, fatigue, practice effects, from his estimate of experimental error. An example of a single group experiment where all of the *S*s underwent the several treatments in the same order is provided by Bruner, Postman, and Mosteller (13). Nine-

TABLE 8

MEAN MAGNITUDE OF CONDITIONED EYELID RESPONSES FOR SUCCESSIVE FIVE-TRIAL BLOCKS DURING EXTINCTION (64)

Sub-ject	Successive Five-Trial Blocks after First 5 Trials				
	2	3	4	5	6
Single Alternation Group					
1	—				—
2	—				.
.	.				.
.	.				.
.	.				.
14	—				—
15	—				
Double Alternation Group					
16	—				—
17	—				—
.	.				.
.	.				.
.	.				.
29	—				—
30	—				
100% Reinforcement Group					
31	—				—
32	—				.
.	.				.
.	.				.
.	.				.
44	—				—
45	—				

Analysis of Variance

Source of Variation	df
Over-all slope	1
Over-all deviations from linearity	3
Between group means	2
Between group slopes	2
Group deviations from estimate	6
Between individual means	42
Between individual slopes	42
Individual deviations from estimate	126
Total	224

TABLE 9
REVERSALS PER SUCCESSIVE ONE-MINUTE INTERVAL OF THE SCHROEDER STAIRCASE UNDER THREE INSTRUCTIONS (13)

Instructions	Minutes	Subjects						
		1	2	...	...	...	18	19
"Alternate" $M=47.4$	1	120	22				42	30
	.							
	.							
	.							
	10	129	11				35	32
"Hold" $M=11.5$	11	19	10				7	5
	.							
	.							
	.							
	20	4	2				1	9
"Natural" $M=21.6$	21	54	20				18	12
	.							
	.							
	.							
	30	25	10	...	...	...	20	10

Analysis of Variance

Source of Variation	df
Subjects (Su)	18
Set (St)	2
Interaction (Su×St)	36
Time-sequence regression	57
Residual sampling variance	456
Total	569

teen Ss were given the task of reversing the Schroeder staircase for successive ten-minute periods under three sets of instructions. Reversals per successive one-minute interval were analyzed with the time-sequence regression lines for the individual Ss being taken out of the total variation of scores as a systematic source of variation. The analysis of variance appeared as in Table 9.

Another complex example, where repeated measurements were analyzed, is found in a study by Lawrence and Miller (89). Their investigation involved a 2×2 factorial design in which individual and group linear regression lines (44) were compared for groups of Ss, doubly classified according to number of reinforced trials and amount of reward. This study appears to be fairly unique in that curvilinear regression lines were also examined.

Comment. Experiments involving the repetitive measurement of a criterion variable for one or more groups of Ss are characteristic of many areas

of psychological research. Similar experiments are not commonly found in experimental agriculture and uses of variance analysis for repeated-measurements designs represent special adaptations made by psychologists. Perhaps the most uniquely psychological of these applications is the situation where the investigator is particularly interested in the comparison of trends, e.g., learning curves, for several independent groups subjected to different experimental conditions. Traditional methods for comparing such trends were largely limited to comparisons of experimental groups either at specified points of the experiment (see Table 1) or with respect to increment or decrement over specified periods (1). Frequently, the experimenter simply compared the over-all means of the several groups, thus completely neglecting the configurations of successive trial means. The analytic methods cited in this section have proven of practical utility for the comparison of group trends, but in general they should be applied with caution since successive measures taken on the same Ss can hardly be regarded as either randomly distributed or independent. There is need for further theoretical work on the topic of repeated measurements, preferably with the aid of mathematical statisticians.

Despite the fact that repeated measurements often appear to exhibit nonlinear trends the writer found that few investigators go to the trouble of fitting curvilinear functions in carrying out analyses of variance. Lindquist (96) and Lewis (92) discuss procedures for testing the goodness of fit of observed successive means to fitted curves in the case of a single group, but there is need for an expository article on the possibilities for comparing curvilinear trends for

independent groups by variance methods.

Finally, it might be noted that the type of investigation in which all Ss undergo several different experimental treatments in the same order should in general be avoided. The experiment by Bruner, Postman, and Mosteller (13) described in Table 9 is a case in point. The design is increased in sensitivity by the separation of individual variations in regression on time from the estimate of error, but this does not overcome the confounding of differences in "set" means with possible temporal effects of fatigue, adaptation, etc. Such confounding of the main experimental factor could have been obviated by randomizing the sequence of experimental conditions among the Ss.

THE LATIN-SQUARE PRINCIPLE OF DESIGN

The fundamental principles of the latin-square design are described in many texts and in articles by Thomson (127), Grant (52), and Edwards (38). The latin square is essentially a triple classification, one variable represented by rows, the second by columns, and the third by treatments which occur once in each row and once in each column. As used by psychologists, the latin-square arrangement has typically been applied as a form of repeated-measurements design where Ss are exposed to several experimental treatments and the investigator desires to take account of the possibility of systematic temporal effects such as transfer, practice, or fatigue.

Single latin-square designs. In the most common design using a single square, the criteria of classification are Ss (rows), trials or successive periods (columns), and experimental treatments (Latin letters). While the latin square is a very compact form of

TABLE 10

NUMBER OF CORRECT RESPONSES MADE BY FIVE SUBJECTS IN READING THE
LUCKIESH-MOSS LOW CONTRAST TEST CHART UNDER
VARIOUS EXPERIMENTAL CONDITIONS (18)

| | *Experimental Conditions** | | | | | |
Subjects	Loud Sound	Weak Sound	Heavy Pressure	Light Pressure	Control	Means
A	21(2)	22(3)	20(4)	22(5)	22(1)	21.4
B	22(4)	16(1)	23(5)	19(2)	23(3)	20.6
C	14(1)	14(5)	23(2)	24(3)	20(4)	19.0
D	29(5)	24(4)	24(3)	24(1)	28(2)	25.8
E	16(3)	15(2)	14(1)	15(4)	13(5)	14.6
Mean	20.4	18.2	20.8	20.8	21.2	

| | *Experimental Days* | | | | |
	1	2	3	4	5
Mean	18.0	21.2	21.8	20.2	20.2

Analysis of Variance

Source of Variation	df
Subjects	4
Days	4
Conditions	4
Residual (error)	12
Total	24

* The entries in parentheses are the days on which the experimental conditions were presented.

design, it should be obvious that its application assumes that interactions among the variables are negligible. If such interactions are present, they are confounded with the other sources of variation and may serve to augment or depreciate the apparent significances of effects. Chapanis, Rouse, and Schachter (18) employed a single 5×5 latin-square design in studying the effects of intersensory stimulation upon contrast sensitivity, as measured by number of correct responses on the Luckiesh-Moss Low Contrast Test Chart. Five Ss were tested under five conditions (loud sound, weak sound, heavy pressure, light pressure, control) on each of five days. The analysis of results followed the form of Table 10.

Leyzorek (91) employed a single 7×7 latin-square design in a study analyzing various types of error scores made in visual interpolation between circular scale markers with differing sizes of scale interval.

Replicated latin-square designs. Studies in which latin-square designs are used may be replicated by employing randomly selected squares of the same size or by applying the same square to several groups of Ss.

In some cases where the number of experimental treatments is small it may be feasible to apply all permutations of order of the several treatments with several Ss undergoing each sequence. The principles involved in these variations have long been used in experimental research and special types of such designs have been variously referred to as permuted double-fatigue orders, balanced orders of presentation, rotation experiments, crossover designs, switchback studies, ABBA orders, etc. The relative advantages and disadvantages of various types of replication of the latin square and methods of analysis are discussed by Grant (52), Edwards (37, 38), Cochran and Cox (27), and Kempthorne (83).

The simplest replicated latin-square design is the 2×2 square in which half of a group of Ss go through two conditions in one order while the remaining Ss go through the conditions in the reversed order (53). Frequently one of the conditions serves as an experimental control. Brogden (12), in a study of sensory conditioning, obtained auditory thresholds from 10 Ss first in the presence of a light stimulus and then in the absence of light, while a second group of 10 Ss was measured in the reversed sequence. Threshold measures were analyzed as in Table 11.

Similar designs were applied by Chernikoff and Brogden (21) and Chernikoff, Gregg, and Brogden (22) in studies of reaction time. In another study employing a 2×2 design of the same kind for a comparison of recall and recognition, the investigators inappropriately interpret the design as a 2×2 factorial (113).

Replication of the same latin-square design for larger squares follows a similar pattern of analysis with the addition of another source

TABLE 11
AUDITORY THRESHOLDS WITH AND WITHOUT LIGHT (12)

	Experimental Group	
Subject*	Threshold with Light	Threshold without Light
1	18.0	20.5
3	20.5	18.0
.	.	.
.	.	.
.	.	.
17	−4.5	5.5
19	15.0	15.0
Subgroup Mean	16.5	18.8
2	20.5	25.5
4	25.5	25.5
.	.	.
.	.	.
.	.	.
18	10.5	10.5
20	10.5	15.5
Subgroup Mean	15.5	18.2
Group Mean	16.0	18.5

Analysis of Variance

Source of Variation	df
Treatment	1
Ordinal position of treatment	1
Sequence of treatment	1
Individual variation of Ss within sequences	18
Error	18
Total	39

* The odd-numbered Ss make up the subgroup for which the threshold with light was made first and the even-numbered Ss are the subgroup for whom the threshold with light was second.

of variation entitled "square uniqueness" by Grant (52) or "latin-square

TABLE 12

Error Scores in Linear Pursuit as a Function of Angle of Arm from Body* (29)

Sequence	Subject	Order						
		I	II	III	IV	V	VI	VII
A	1	(180°)	(210°)	(240°)	(270°)	(300°)	(330°)	(360°)
	2	—	—	—	—	—	—	—
	3	—	—	—	—	—	—	—
	4	—	—	—	—	—	—	—
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
G	25	(210°)	(240°)	(270°)	(300°)	(330°)	(360°)	(180°)
	26	—	—	—	—	—	—	—
	27	—	—	—	—	—	—	—
	28	—	—	—	—	—	—	—

Analysis of Variance

Source of Variation	df
Angle	6
Sequence of angles	6
Ordinal position of angles	6
Individual differences within sequences	21
Square uniqueness	30
Remainder	126
Total	195

* Data in body of table not provided; entries indicate sequence of angles; each score was the mean of 20 trials at a given angle.

error" by Edwards (38). The latter author emphasizes application of tests of homogeneity of variance before pooling terms for an over-all error estimate. A 7×7 replicated latin-square design employing the same square for several groups of Ss was used by Corrigan and Brogden (29) in studying the effect of bodily angle upon precision of linear pursuit movements. Twenty-eight Ss were randomly allocated to seven groups of four Ss each. Each group went through the pursuit task at seven angles with varying orders of presentation in a latin-square design. Error scores, based on combining 20 trials at each angle, were analyzed as in Table 12.

A similar analysis of an experimental design involving replication of the same 24×24 latin square with two Ss in each sequence was carried out by Corrigan and Brogden (30). Two studies by Gregg and Brogden (66, 67) also involved the application of replicated 6×6 latin squares.

As indicated above, when the number of experimental treatments is small it is possible to utilize the latin-square principle by providing for every possible permutation of

order of presentation. This arrangement and the method of analysis are discussed by Grant (52). Ryan, Cottrell, and Bitterman (118) employed such a design in a study of muscular tension when they assigned four *S*s to each of the six possible orders of three experimental conditions (glare, noise, control), but their analysis of results was limited to treating the design as a double classification without replication. In an extension of the device of using every permutation of experimental orders, Grant, Jones, and Tallantis (62), studying concept formation by means of a card-sorting experiment, employed what might be called a double latin-square design since each group of *S*s repeated their assigned order of three experimental treatments. Under the conditions set up by the investigators, there were four *S*s for each of the 24 possible permutations specified.

The other major replicated latin-square design involves the application of several different randomly selected squares. No examples of this design other than those described by Grant (52) and Edwards (38) were found in the literature surveyed.

Combined latin-square and factorial design. The latin-square principle may be combined with a factorial arrangement of treatments in various ways (37, 52, 27). One simple method, for example, is to have a 4×4 square, with rows representing *S*s and columns representing successive trials, in which the four latin treatments are A_1B_1, A_1B_2, A_2B_1, and A_2B_2. This specific design was used by Prentice (115) in a study of the relation of distance to apparent size of figural after-effects. Four *S*s were subjected to the four different treatment combinations on four successive days in a latin-square design. Each treatment was the combination of one of

TABLE 13

POINTS OF SUBJECTIVE EQUALITY FOR EACH SUBJECT FOR DIFFERENT DISTANCES AND CONDITIONS OF SATIATION* (115)

Subject	Day			
	1	2	3	4
A	2S	6NS	2NS	6S
B	2NS	6S	2S	6NS
C	6S	2S	6NS	2NS
D	6NS	2NS	6S	2S

Analysis of Variance

Source of Variation	df
Days	3
Subjects (sequences)	3
Satiation vs. no satiation	1
Distance	1
Interaction: Satiation × Distance	1
Remainder	6
Total	15

* Data not provided. The numbers in the body of table refer to distance (2 m. or 6 m.) at which *S* made his judgment; the letters S and NS refer to the conditions of "satiation" and "no satiation."

two distances of stimulus and one of two conditions of "satiation." Analysis was carried out for only three sources of variation, viz., one *df* for the two variations of distance, seven (?) for individuals, and seven for error, but might have been extended as shown in Table 13.

Another variation of the principle of combining latin-square and factorial design was found in a study of delayed response performance by Meyer, Harlow, and Settlage (104). In this experiment measures from sets of four monkeys were arranged in a 4×4 latin square, with four learning periods (rows) and four experimental conditions (columns). Within each of the 16 cells of the square there were 16 entries for all

combinations of four types of object pairs and four lengths of delay. Separate analyses were carried out for normal, unilateral damaged, and frontal damaged Ss.

A third variation of combining latin-square and factorial principles in a replicated design was used by Cameron and Magaret (16) in a study of responses to incomplete sentences. In this study a 2×4 design was employed, each subcell indicating the combination of two factors. Replication of sequences by having seven Ss undergo each order of conditions allowed for the differentiation of sequence variation from variation among Ss in the same sequence.

A still more complex design involving latin-square and factorial principles in replication was used by Postman and Bruner (114) in a study of the relation of set and perceptual behavior. Their analysis, however, is difficult to explicate since composite sources of variation were dubiously partitioned and an attempt was made to analyze confounded interactions.

Greco-latin square designs. A further extension of the latin-square principle is to add another experimental treatment to the latin-square design in such a way that each new treatment appears but once with each Latin letter treatment (44, 45, 52). No "pure" examples of greco-latin square designs were found in the literature surveyed.

Comment. In a recent article McNemar (102) has discussed a type of application of the latin-square design which has been neglected by psychologists. As noted above, the common use of the latin square has been the case where one classification of the data consists of experimental treatments while the other two classifications consist of uncontrollable sources of variation, e.g., Ss and

trials. McNemar suggests the use of latin square as an economical form of three-factor design, when each factor consists of the same number of levels, and the mixed design where two of the three classifications are experimental factors. An example of the latter case is provided by Garrett and Zubin (49) who describe a study of color recognition by the dark-adapted eye where the three classifications in a 4×4 latin square were order of presentation (rows), levels of illumination (columns), and color (Latin letters). If the rows in this study had represented, say, four levels of dark adaptation, instead of order of presentation, this study could have served as an example of a three-factor latin-square design.

Of perhaps more importance is McNemar's contention that the latin-square design is rarely applicable in psychological research because the basic assumption of negligible interactions among the three classification variables is generally violated, especially in the design where Ss form one of the criteria of classification. McNemar concludes that the use of the latin square is "defensible only in those rare instances when one has sound a priori reasons for believing that the interactions are zero" (102, p. 400).

There is no question but that the standard mathematical model of the latin-square design assumes that interactions are negligible and that statistical inference is most dependable when this assumption holds. But the writer does not agree with McNemar when he states that too many "significant" F's are obtained when this assumption is not met because the residual term, containing both interaction and the ordinary error, tends to be smaller than the interaction properly used in the denominator for F. In the first place, it is clear that

the single latin square never provides an estimate of "pure" error of the type available when replication within the same subclasses is carried out. The residual of the latin square is always an admixture of confounded first-order and second-order interactions. This admixture, provided that the interactions are negligible, furnishes an unbiased estimate of "pure" error. When "significant" (but untestable within the design) interactions are present the residual could possibly be reduced in a specific experiment if the interaction(s) happened to follow the pattern of one or more of the three major classifications, but in the long run, if squares were always randomly selected, it would be expected that significant interaction(s) would tend to increase the residual, thus inflating the estimate of error. In such a case, if the experimenter were interested in testing the significance of a main effect against an estimate of "pure" error (regardless of the presence of interactions), there would be no such estimate available, and his purpose could not be met. If, unknowingly, a main effect were tested against such an inflated estimate of "pure" error, the *F* ratio would tend to be too small.

Let us assume, with McNemar, however, that the experimenter is desirous of testing the significance of main effects over and above the presence of possibly significant interactions. This would be analogous to testing main effects against interaction in the two-factor replicated design. The mathematical model for the two-factor case is simple because only one interaction is present. The important point, however, about the two-factor design is that the observed interaction term is assumed to be composed of two additive components, interaction variance plus error variance, while an observed main ef-

fect is assumed to be composed of three additive components, main effect variance plus interaction variance plus error variance. Since the residual term in the latin square is made up of several confounded interactions, it is impossible to set up a simple "components-of-variance" model (see below) as for the two-factor design. Nevertheless, for practical purposes one can assume that the observed residual of the latin square is composed of two components: confounded interactions variance plus error variance. Each observed main effect would then be assumed to consist of three components: main effect variance plus confounded interactions variance plus error variance. The consequent *F* ratio of main effect mean square over residual mean square should then tend to give an unbiased test of the significance of main effect over and above the presence of significant interactions. McNemar's contention that the denominator of the *F* test should properly consist of interaction alone, i.e., separated from error variance, has, so far as the writer is aware, no precedent in analysis of variance methods. In any case, however, it is clear that the presence of significant interactions negates the application of the usual latin-square design. No mathematical justification is readily available for inference in the case where some of the interactions are "significant."

DESIGNS INVOLVING ANALYSIS OF COVARIANCE

In some investigations designed for analysis of variance it may not be feasible to control or classify the data on the basis of one or more relevant variables which can, however, be measured. The addition of covariance analysis (44) to the experimental design allows for adjustments to be made in experimental comparisons

on the basis of the regressions of the variable of primary importance on these other relevant variates. Covariance analysis may be carried out for all of the experimental designs so far presented. Discussions of covariance are readily available both for single-classification designs (36, 37, 76, 95, 101, 124) and for multiple-classification designs (37, 76, 95, 124), as well as for the case of one independent or control variable (36, 37, 76, 95, 101, 124) or two independent variables (76, 124). Snedecor (124) also provides an example of covariance in a latin-square design.

Applications of covariance analysis in experimental design were not too common in the literature surveyed. In general, moreover, when covariance was used little descriptive detail was provided. Bernberg (9) adjusted error scores for three groups of rats, learning a maze under three different conditions, on the basis of differential food intake in a single-classification design. In a study of reminiscence, Buxton and Bakan (15) adjusted criterion scores based on differences between rest and no-rest conditions by "correction" for recall trial difference scores. Buxton and Ross (14) similarly applied covariance analysis to a two-factor design in a study of the relationship between reminiscence and type of learning technique.[4] Reynolds (117), in a study of resistance to extinction, considered covariance adjustments of learning scores on the basis of scores on a previously trained habit, but rejected the plan because of heterogeneity of variances and low correla-

tions. He then adjusted for training time in an analysis of the extinction scores. Glixman (51) applied covariance analysis to a $3 \times 3 \times 2$ factorial design in a study of recall of completed-incompleted tasks under differing conditions of stress. Covariance adjustments were made for scores based on number of incompleted-recalled tasks in terms of total number of incompleted tasks.

The complex problem of handling disproportionate subclass frequencies in a double-classification 2×4 factorial design with covariance analysis is exemplified in a study by Fitch, Drucker, and Norton (46), who used a procedure developed by Tsao (128). A rather full explanation of design, basic assumptions, and analytic procedures is provided by this study.

Comment. As noted by Fisher (45) and others, analysis of covariance in experimental design may be used for two major purposes: (*a*) to increase the precision of experimental comparisons by statistically controlling for sources of variation which do not lend themselves to experimental control, and (*b*) to aid in the interpretation of the results of an experiment. In the former case the experimenter should be sure that the "control" variable is independent of treatment effects (45). Ordinarily he is not particularly interested in studying the relationship between the concomitant measures and the primary variable being investigated (95). The standard example of a supplementary variable which can frequently be employed to improve the precision of an experiment is pretest scores on the same kind of performance which is to be measured during the experiment itself.

More care must be taken in deciding to utilize analysis of covariance for the second purpose. In this case supplementary measures are ordi-

---

[4] The main rationale for using covariance in this study was to "remove" variance due to using the same *S*s under experimental and control conditions. Grant (52) illustrates how this same study might have been analyzed by interpreting the arrangement as a 2×2 greco-latin square.

narily taken during the course of the experiment and hence variations in the concomitant variable may be a function of the experimental treatments. The obvious difficulty in applying analysis of covariance here is that adjustment of the primary variable may remove part of the treatment effect itself. The experimenter may wish, however, to find out whether there are significant differences in treatment effects on the primary variable when the secondary variable is "equalized" over all groups. In such cases it is generally profitable to carry out not only an analysis of covariance to "eliminate" possible effects of the secondary variable, but also separate analyses of variance of both the secondary variable and the unadjusted primary measures, as well as careful examination of regression and correlation coefficients. Comparison of the several analyses will tend to clarify the extent to which experimental effects on the primary variable act directly or indirectly through the mediation of the concomitant variable or covariate. For example, to paraphrase Snedecor (124, p. 335), in the study by Glixman (51) cited above: Did the Ss show differences in number of incompleted-recalled tasks under varying conditions of stress because of differences in total number of incompleted tasks, or in spite of them? Excellent discussions of the use of analysis of covariance to improve understanding of experimental structure are presented by Edwards (37) and others (27, 83, 124).

In addition to the complexities of procedure and interpretation which generally arise when analysis of covariance is applied to designs involving multiple classification, or when there is more than one concomitant variable, other difficulties sometimes arise in the use of covariance analysis. At times the regression of primary variable upon supplementary variable may be nonlinear, necessitating the adjustment of primary variable on the basis of curvilinear regression (80). In other cases the experimenter may discover that there are problems in the choice of appropriate regression coefficients for estimation of the main variable because of marked heterogeneity of regression from subclass to subclass. Jackson (73, 74) provides a detailed discussion of such problems and others and suggests possible solutions.

In the context of analysis of covariance, special mention should be made of the Johnson-Neyman technique (78, 81). As noted above, one of the major uses of covariance methods is to adjust experimental comparisons for extraneous causes of variation. Frequently, such adjustment is designed to "equate" experimental groups when it is not feasible to increase precision by pairing cases or otherwise matching the several groups with respect to the relevant measures. The Johnson-Neyman technique not only furnishes a test of whether a statistically significant difference exists between the means of the groups being compared but, in addition, specifies the range of control variables for which a conclusion of significant difference may be regarded to hold. Moreover, no special difficulties are involved when the groups are unequal in number. Johnson and Fay (81) provide the detailed computational and graphical solution for a problem in which the social studies achievement of 90 pupils who excel in the ability to predict the outcome of given events is compared with the social studies achievement of 90 pupils who are poor predictors. The null hypothesis rejected on the basis of the analysis was that no difference exists in mean achievement

between superior and inferior predictors when the effects of chronological and mental ages are controlled. The unique surplus information contributed by the Johnson-Neyman technique indicated the range of mental age and chronological age for which the conclusion of significant difference was valid.

General Considerations

Models, assumptions, and transformations in analysis of variance. Before valid inferences may be drawn from an analysis of variance, the data must reasonably satisfy certain assumptions made about the underlying mathematical models used in the analysis and subsequent tests of significance. In recent years various sets of assumptions have been proposed about the elements in the linear models whereby analysis of variance is used for statistical inference. Pointing out that Fisher (44) had originally introduced the twofold conception, Eisenhart (41), in 1947, elaborated on the viewpoint that analysis of variance involves one of two basic models, each appropriate for the solution of a different class of problems: *Model I* to detect or estimate fixed relations among population means, and *Model II* to detect or estimate components of random variation ascribable to the different factors being investigated. The former is frequently referred to as the *standard* model while the latter is commonly called the *components-of-variance* model (105).

In brief the major distinction between the two models is that Model I assumes that treatment and other designated effects are additive fixed constants, introducing systematic variation, while Model II assumes that treatment and other effects are random variables each having a normal distribution. Both models assume that experimental (residual) errors are independently and normally distributed with a constant variance. The decision as to whether a given element in the linear model is best represented by a *mean*, indicating a systematic source of variation, or by a *variance*, indicating a random source of variation, depends upon the extent to which the respective variable was randomly sampled. In many experiments some of the effects are best regarded as fixed, e.g., the usual case for experimental treatments which are rarely randomly drawn from a population of possible treatments, while other effects may be regarded as introducing random variation, e.g., effects assignable to *Ss* drawn at random from a specified population. When both types of elements are present, the underlying model is described as "mixed."[5]

The majority of published psychological studies employing variance designs give little evidence that investigators pay much attention to the several assumptions underlying analysis of variance. For instance, although tests of homogeneity of subgroup variance are readily available, e.g., Bartlett's test (37, 76, 84, 124), the L_1 test (76, 84), the M test (76, 84), and Box's test (36), the assumption that experimental errors have

---

[5] A recent review by Crump (31) indicates that Eisenhart's so-called Model I, Model II, and Mixed Model have been supplemented by Tukey with Models III, IV, V, and X, all involving somewhat different assumptions. Many of the analyses described by Kempthorne (83) are based upon finite "randomization" models, involving no assumption about normality of error distributions. Other models have been proposed which are also nonparametric, i.e., make no assumption about the form of the population distributions (105, 106). In general, the assumptions involved in all of these models are less restrictive than the usual set of assumptions, thus broadening the potential applicability of analysis of variance techniques.

equal variance appears to be tested only in a minority of experiments. The assumptions that errors are uncorrelated, with constant variance, appear on both empirical and theoretical grounds to be somewhat more critical than the assumption of normality of distribution of the errors (25, 27, 83).

Cochran, in a detailed discussion of the consequences when the assumptions for analysis of variance as a technique for carrying out tests of significance of differences among means are not satisfied, states that "the principal methods for an improved analysis are omission of certain observations, treatments, or replicates, subdivision of the error variance, and transformation to another scale before analysis" (25, p. 37). The method most frequently resorted to by psychologists who take cognizance of violation of assumptions is transformation of the scale. The rationale and conditions for various kinds of transformations are most fully discussed by Bartlett (5). Briefer accounts are available in other sources (37, 76, 83, 124). Such transformations are frequently intended to stabilize error variance, especially in cases where variances within the subclasses show a functional relationship with subclass means.

In the experimental literature, Haggard (68, 69) provides a careful investigation of the problem of selecting proper measures of GSR data for analysis of variance procedures and the effects of using inappropriate measures. Among the specific transformations that were utilized in recent psychological studies were square-root transformation of number of reversals of perspective (13), log transformation of latency scores (4, 93), log transformation of hoarding scores (72), log transformation of

number of contacts in a pursuit-meter task (30), reciprocal transformation of latency scores (89), arcsine transformation of percentages (75, 94, 109), and transformation of obtained scores to per cent of prestimulus values (35).

The F test in analysis of variance. In the single-classification design the error term of the F ratio is provided by the "within-groups" mean square. Similarly for the double-classification design with a single observation in each cell, the denominator of the F ratio is furnished by the interaction or remainder mean square. Complications arise, however, in selecting the proper denominators for the F ratios in the case of the double-classification design with several observations in each subclass. In general, the most widely used procedure has depended upon whether or not the investigator decides that the interaction F test is "significant." If the interaction F test is not found to be significant, the investigator may either use the "within-cells" mean square as his error estimate in testing the significance of variation among main effects or pool the "interaction" and "within-cells" SS in arriving at an error estimate. If, however, he finds that the interaction mean square is significant, he generally employs the latter square in his F tests for the main criteria of classification. Strictly speaking, the use of the interaction mean square as an estimate of error in this case involves application of a "components-of-variance" model to the data, since it is thereby assumed that the expected value of the mean square for an apparent main effect is made up of a linear combination of error variance, interaction variance, and the intrinsic main effect variance itself. The hypothesis is thus being tested that the intrinsic main effect is not

significant over and above any variation attributable to both random error and the effect of interaction.

With higher order multiple-classification or factorial designs, the selection of appropriate F tests becomes more complex. Again, psychologists have generally proceeded "from the bottom up" in setting up F ratios from an analysis of variance table. For example, if a three-factor design with several replications per subclass is involved, the first test is generally made by setting up the F ratio of the highest order interaction mean square over the "within-cells" mean square. If this highest order interaction is found to be "not significant," it may then be pooled with the "within-cells" term to provide the denominator for F tests of the next highest interaction terms. On the other hand, if the highest order interaction is found to be "significant," the F tests of the next highest interaction mean squares are made with the highest order interaction mean square in the denominators. When the investigator arrives at the main variables of classification he generally has used as his denominator for F ratios the pooled interactions and residual, if none of the preceding F's has been significant, or the interaction mean square of largest magnitude which contains the elements assumed to be contributing to the apparent variability of a given main variable. This *ad hoc*, somewhat intuitive procedure for arriving at F ratios may frequently be criticized from the standpoint of the variance components assumed to be operating in the specific situation or because the elements assumed to be random variables in the linear model may more logically be assumed to represent fixed parameters. If, however, a "components-of-variance" model (41) is justified by the data and sampling methods used in the study, a more appropriate method of testing relations is available. Among others (31, 83, 84, 105), Cochran (26) has recently discussed in detail the problems arising in testing a null hypothesis about several means when an appropriate denominator for an F ratio is not immediately provided by the "expected mean squares" in the analysis of variance table. The procedure suggested involves setting up what Cochran calls an F' test where numerator and denominator of the F ratio are linear combinations of mean-square terms arranged in such a way that the treatment effect to be tested is present only in the numerator, while all remaining assumed components of variance are present in both the numerator and denominator. The respective df's for the composite ratio are determined according to approximations proposed by Satterthwaite (119, 120). The formulation of such ratios is facilitated by the provision of expected mean squares for many commonly used experimental designs by Snedecor (124) and Mood (105).

Individual tests of significance in the analysis of variance. Investigators frequently desire to follow an over-all analysis of variance with tests of the significances of differences between individual pairs or groups of treatment means. When the F associated with a given classification is found to be significant, the most commonly used procedure has been the method described by Lindquist (95), in which t tests are applied to the selected means, using standard errors of differences based on the appropriate error variance from the analysis of variance. Confidence intervals can be set up on the same basis. When the F ratio representing a given classification is found to be not significant, the investigator generally ceases his anal-

ysis. In recent years this simple alternative operation has been criticized and further extensions of analysis of variance have been suggested. Dixon and Massey (36) propose a "test for extreme mean" applicable in the situation where one group of Ss is a control group, while the remaining groups are experimental groups. Johnson (76), utilizing a suggestion made by Fisher (45), discusses how *selected* pairs of means may be compared by lowering the p level for significance in accordance with the possible number of comparisons. Snedecor (124) and Cochran and Cox (27) warn about the dangers of testing differences suggested by the data and present methods for subdividing the treatment and error SS for relevant individual and group comparisons.

Subdivision of treatment SS is especially applicable in experiments where the levels of a given factor represent varying amounts or categories along a treatment continuum, e.g., degrees of learning. Kelman (82), for example, had four groups of Ss (Control, Success, Failure, Ambiguous) which furnished three orthogonal comparisons: (*a*) Control group vs. experimental groups (3C-S-F-A); (*b*) Ambiguous reinforcement vs. clearcut Success or Failure (2A-S-F); and (*c*) Success vs. Failure (S-F). Johnson and Tsao (79), in a $4 \times 7 \times 2 \times 2$ factorial study dealing with the determination of differential limen values, furnish a detailed discussion of the application of orthogonal polynomials (44) in expressing the relationships between the factors, e.g., weight, rate, etc., and the limen values. The procedure of fitting orthogonal polynomials to a given factorial classification with associated tests of significance for linear regression, parabolic regression, etc. can frequently be used to furnish infor-

mation and answers to questions which are not supplied by the overall F test for a given set of treatment means (27, 124).

Perhaps the most simple and practical procedure for comparing individual means now available to the investigator who is not satisfied with the results of an over-all F test is that presented by Tukey (129). In this procedure, after finding a significant F for a set of treatment means, one successively applies a "gap" test, a "straggler" test, and a new F test to subgroups among the treatment means to detect distinguishable groups.

The power of analysis of variance tests. Almost universally, the psychologist has limited his attention in testing hypotheses to consideration of errors *of the first kind* (Type I errors), i.e., rejecting hypotheses when they are true. The risk of committing errors *of the second kind* (Type II errors), i.e., accepting false hypotheses, has in general entered very little into his schema of statistical inference. The usual test of significance involves the specification of a so-called critical region which controls only the risk of error of the first kind. Thus, for example, if the critical region is set at the .05 level for an F test, the experimenter is in effect declaring that rejection of the null hypothesis when p equals or exceeds this value will be wrong only 5 per cent of the time.

What happens, however, when a null hypothesis is not rejected, i.e., if the F ratio is smaller than, say, the tabulated value for the .05 level? Most investigators appreciate in theory that such a finding does not mean that the null hypothesis is proved. And yet the tendency is strong to accept the null hypothesis and draw the conclusion that no differences among means are present. The *power function* of a given test of significance is

designed to indicate the probabilities of rejecting a specified hypothesis when alternative hypotheses are assumed to be true. In the usual case where the specified hypothesis is a null hypothesis, i.e., all of a group of means are equal, the probability of rejecting the null hypothesis when the true means are in fact different depends upon the significance level selected for the test, the magnitude of the differences among the means, the size of the error variance, and the number of replicates.

Two major approaches have been devised for determining the power of analysis of variance tests. In the older method developed by Tang (126), the alternative hypothesis to the null hypothesis is expressed in terms of the variance of a finite set of assumed population means equal in number to the number of observed means involved in the F test. In a more recent approach described by Ferris, Grubbs, and Weaver (43), the alternative hypothesis is expressed in terms of a set of normally distributed population means, these means representing a sample from a normal superpopulation with variance bearing a specified ratio to the error variance involved in the F test. Since, however, this latter paper presents a somewhat limited set of curves for estimating the power of the analysis of variance test at only the .05 significance level, further discussion will be limited to the Tang approach.[6]

The method developed by Tang assumes that the observations can be expressed in terms of Model I (see above), which assumes a linear combination of mean effects and errors which are normally and independently distributed with constant variance. Tang presents fairly extensive tables for varying pairs of df, under the assumption that either a .05 or .01 level of significance is being employed for rejection of the null hypothesis. In these tables the probabilities of error of the second kind are indicated for varying sizes of ϕ, a variance ratio with numerator derived from the assumed alternative hypothesis. Lehmer (90) subsequently prepared tables providing the value of ϕ required for a specified probability of error of the second kind. Tang's tables are reproduced with extensive discussion of their use in Kempthorne (83) and Mann (99), while Lehmer's tables for Type II errors of probability .3 and .2 are available in Dixon and Massey (36).

The power function of analysis of variance tests is very useful (a) for estimating the sample sizes that will reasonably guarantee a desired probability of error of the second kind for a specified alternative hypothesis and a designated level of significance, and (b) for determining the power of a test, if the sample sizes and significance level have already been fixed. Perhaps the chief implication of the power concept in relation to current psychological research is the conclusion that many experiments are carried out without sufficient replication to insure a reasonable chance of detecting experimentally important differences in treatment means. Kempthorne (83), for example, estimates that six replicates in each subclass are necessary in a $2 \times 2 \times 2$ factorial experiment to insure with a probability of .95 that a true difference of the two means for a given factor equal in size to the error standard deviation will

[6] Actually the paper of Ferris, Grubbs, and Weaver (43) presents "operating-characteristics" curves which are related to power curves as x is to $1-x$, i.e., complementary. Eisenhart et al. (42) also provide a brief discussion of operating-characteristic functions for analysis of variance tests based upon Eisenhart's Model II.

be detected (using a .05 test of significance).

The power conception is somewhat contrary to the commonly accepted notion that an experimenter should insist on a very stringent level of significance before rejecting a null hypothesis when his samples are relatively small. Such a notion, paradoxically defended on the grounds of conservatism, has probably resulted in the premature dismissal of many potentially important areas of experimentation. It might be a worthwhile addendum to current methodology if many exploratory, small-sample experiments were primarily devoted, not to tests of hypotheses, but to obtaining estimates of error variance. Once such estimates of error variance have been obtained, the experimenter is in a position to determine the sample sizes necessary to detect differences between treatments regarded to be of practical or theoretical significance.

An overview of psychological research design and analysis. Because of the widely accepted thesis that experimental design and statistical analysis are "dynamic" aspects of the same research "whole," and inasmuch as many statistical and measurement techniques, e.g., regression and correlational analysis, the *t* test, chi square, discriminant functions, etc., can be subsumed under analysis of variance and the *F* distribution, a thorough survey of variance designs used by psychologists could have been extended far beyond the limits of the present article. Minimum attention, moreover, has been given to the general theory of experimental design in this survey. Basic concepts such as experimental control, statistical control, randomization, replication, balance, efficiency, precision, orthogonality, comprehensiveness,

self-containedness, etc. enter in the adoption of any specific experimental design, but, in general, the journal reporting of studies is not amenable to elaboration of underlying principles.

With regard to specific experimental arrangement, cursory examination of books which have treated experimental design in agricultural, biological, or industrial research, from Fisher's classic (45) to the recent comprehensive presentations by Cochran and Cox (27) and Kempthorne (83), reveals that psychologists have generally utilized only the simpler and "complete" experimental configurations. Of the 150 experimental plans presented by Cochran and Cox (27), only a small minority seem to have appeared in psychological research design. As noted previously, various devices of *deliberate confounding* or *partial confounding*, especially applicable in higher-order factorial studies, seem to be only rarely considered, despite their early introduction into the psychological literature by Baxter (6) and their potential experimental and practical advantages. This is not to say that methodology from one area of research can be routinely applied in another area, but the increasing frequency with which variance designs have been applied in psychological research probably indicates a trend which can be expected to continue.

The practicing researcher, of course, finds difficulties in keeping up with current developments and refinements in the area of experimental design and analysis. A real service is being performed by the excellent summaries being presented in the *Annual Review of Psychology* (39, 54, 103). Nevertheless, as pointed out by Johnson (77) in a recent discussion of the contribution of statistical science to educational and psychological re-

search, the "newer developments in the field have mainly been specialized devices for specialized purposes" with three basic principles of experimentation—replication, randomization, and control of variability—being the foundation stones of modern experimental design. Current psychological research, as we have seen, has been tremendously influenced by the "Fisherian revolution in methods of experimentation" (133).

SUMMARY

This article has presented a survey of the major types of experimental design involving analysis of variance which have characterized psychological research during recent years. The survey is implemented by brief reference to specific studies utilizing a variety of experimental configurations which have appeared in the literature. Some comments were made about the appropriateness of design or analysis in particular instances, followed by a discussion of general considerations in application of variance design and analysis.

BIBLIOGRAPHY

1. ALEXANDER, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. AMMONS, R. B. Effects of pre-practice activities on rotary pursuit performance. *J. exp. Psychol.*, 1951, **41**, 187–191.
3. ANDERSON, R. L. Missing-plot techniques. *Biometrics*, 1946, **2**, 41–47.
4. ARNOLD, W. J. Simple reaction chains and their integration. II. Heterogeneous chaining with terminal reinforcement. *J. comp. physiol. Psychol.*, 1947, **40**, 427–440.
5. BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39–52.
6. BAXTER, B. Problems in the planning of psychological experiments. *Amer. J. Psychol.*, 1941, **54**, 270–280.
7. BEEBE-CENTER, J. G., BLACK, P., HOFFMAN, A. C., & WADE, M. Relative per diem consumption as a measure of preference in the rat. *J. comp. physiol. Psychol.*, 1948, **41**, 239–251.
8. BENDIG, A. W., & BRAUN, H. W. The influence of the genotype on the retention of maze habit in the rat following electroshock convulsions. *J. comp. physiol. Psychol.*, 1951, **44**, 112–117.
9. BERNBERG, R. E. A comparison of the effects of electroconvulsive shock and electronarcosis upon the learning ability of young rats. *J. comp. physiol. Psychol.*, 1951, **44**, 50–60.
10. BLACK, J. W. A compensatory effect in vocal responses to stimuli of low intensity. *J. exp. Psychol.*, 1950, **40**, 396–397.
11. BRAY, D. W. The prediction of behavior from two attitude scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 64–84.
12. BROGDEN, W. J. Sensory conditioning measured by the facilitation of auditory acuity. *J. exp. Psychol.*, 1950, **40**, 512–519.
13. BRUNER, J. S., POSTMAN, L., & MOSTELLER, F. A note on the measurement of reversals of perspective. *Psychometrika*, 1950, **15**, 63–72.
14. BUXTON, C. E., & ROSS, H. V. Relationship between reminiscence and type of learning technique in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 41–46.
15. BUXTON, C. E., & BAKAN, M. B. Correction vs. non-correction learning techniques as related to reminiscence in serial anticipation learning. *J. exp. Psychol.*, 1949, **39**, 338–341.
16. CAMERON, N., & MAGARET, ANN. Experimental studies in thinking: I. Scattered speech in the responses of normal subjects to incomplete sentences. *J. exp. Psychol.*, 1949, **39**, 617–627.
17. CARPENTER, A. The rate of blinking during prolonged visual search. *J. exp. Psychol.*, 1948, **38**, 587–591.
18. CHAPANIS, A., ROUSE, R. O., & SCHACHTER, S. The effect of inter-sensory stimulation on dark adaptation and night vision. *J. exp. Psychol.*, 1949, **39**, 425–437.
19. CHAPANIS, A., & LEYZOREK, M. Accuracy of visual interpolation between scale markers as a function of the num-

ber assigned to the scale. *J. exp. Psychol.*, 1950, **40**, 655–667.

20. CHERNIKOFF, R., & BROGDEN, W. J. The effect of instructions upon sensory pre-conditioning of human subjects. *J. exp. Psychol.*, 1949, **39**, 200–207.

21. CHERNIKOFF, R., & BROGDEN, W. J. The effect of response termination of the stimulus upon reaction time. *J. comp. physiol. Psychol.*, 1949, **42**, 357–364.

22. CHERNIKOFF, R., GREGG, L. W., & BROGDEN, W. J. The effect of fixed duration stimulus magnitude upon reaction time to a response terminated stimulus. *J. comp. physiol. Psychol.*, 1950, **43**, 123–128.

23. CHILD, I. L., & GROSSLIGHT, J. H. The effect of substitute activity as depending upon the nature of the similarity between substitute and original activity. *Amer. J. Psychol.*, 1947, **60**, 226–239.

24. CITRON, A. F., CHEIN, I., & HARDING, J. Anti-minority remarks: a problem for action research. *J. abnorm. soc. Psychol.*, 1950, **45**, 99–126.

25. COCHRAN, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, **3**, 22–38.

26. COCHRAN, W. G. Testing a linear relation among variances. *Biometrics*, 1951, **7**, 17–32.

27. COCHRAN, W. G., & COX, G. M. *Experimental designs.* New York: Wiley, 1950.

28. CONKLIN, J. E. Three factors affecting the general level of electrical skin resistance. *Amer. J. Psychol.*, 1951, **64**, 78–86.

29. CORRIGAN, R. E., & BROGDEN, W. J. The effect of angle upon precision of linear pursuit movements. *Amer. J. Psychol.*, 1948, **61**, 502–510.

30. CORRIGAN, R. E., & BROGDEN, W. J. The trigonometric relationship of precision and angle of linear pursuit movements. *Amer. J. Psychol.*, 1949, **62**, 90–98.

31. CRUMP, S. L. The present status of variance component analysis. *Biometrics*, 1951, **7**, 1–16.

32. CRUTCHFIELD, R. S. Efficient factorial design and analysis of variance illustrated in psychological experimentation. *J. Psychol.*, 1938, **5**, 339–346.

33. CRUTCHFIELD, R. S. The determiners of energy expenditure in string-pulling by the rat. *J. Psychol.*, 1939, **7**, 163–178.

34. CRUTCHFIELD, R. S., & TOLMAN, E. C. Multiple-variable design for experiments involving interaction of behavior. *Psychol. Rev.*, 1940, **47**, 38–4

35. DAVIS, R. C. Motor responses to auditory stimuli above and below threshold. *J. exp. Psychol.*, 1950, **40**, 10 120.

36. DIXON, W. J., & MASSEY, F. J. *Introduction to statistical analysis.* New York McGraw-Hill, 1951.

37. EDWARDS, A. L. *Experimental design psychological research.* New York Rinehart, 1950.

38. EDWARDS, A. L. Homogeneity of variance and the Latin square design. *Psychol. Bull.*, 1950, **47**, 118–129.

39. EDWARDS, A. L. Statistical theory a research design. In C. P. Stone (Ed. *Annual review of psychology.* Stanford Annual Reviews, 1951. Pp. 335–35

40. EDWARDS, A. L., & HORST, P. The calculation of sums of squares for interactions in the analysis of variance. *Psychometrika*, 1950, **15**, 17–24.

41. EISENHART, C. The assumptions underlying the analysis of variance. *Biometrics*, 1947, **3**, 1–21.

42. EISENHART, C., HASTAY, M. W., WALLIS, W. A. *Selected techniques statistical analysis.* New York McGraw-Hill, 1947.

43. FERRIS, C. D., GRUBBS, F. E., WEAVER, C. L. Operating characteristics for the common statistical tests significance. *Ann. math. Statist.*, 194 **17**, 178–197.

44. FISHER, R. A. *Statistical methods for research workers.* Edinburgh: Oliver and Boyd, 1925–1946; New York: Hafner 1950.

45. FISHER, R. A. *The design of experiments.* Edinburgh: Oliver and Boyd, 193 1951.

46. FITCH, M. L., DRUCKER, A. J., & NORTON, J. A. Frequent testing as a motivating factor in large lecture classes. *J. educ. Psychol.*, 1951, **42**, 1–20.

47. FRANKLIN, J. D., & BROŽEK, J. The relation between distribution of practice and learning efficiency in psychomotor performance. *J. exp. Psychol.*, 194 **37**, 16–24.

48. FURCHTGOTT, E. Effects of total body X-irradiation in learning: an exploratory study. *J. comp. physiol. Psychol.*, 1951, **44**, 197–203.

49. GARRETT, H. E., & ZUBIN, J. The analysis of variance in psychological research. *Psychol. Bull.*, 1943, **40**, 23 267.

50. GEBHARD, M. E. The effect of success and failure upon the attractiveness of activities as a function of experience, expectation, and need. *J. exp. Psychol.*, 1948, 38, 371–388.

51. GLIXMAN, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, 39, 281–295.

52. GRANT, D. A. The Latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, 45, 427–442.

53. GRANT, D. A. The statistical analysis of a frequent experimental design. *Amer. J. Psychol.*, 1949, 62, 119–122.

54. GRANT, D. A. Statistical theory and research design. In C. P. Stone (Ed.), *Annual review of psychology.* Stanford: Annual Reviews, 1950. Pp. 277–296.

55. GRANT, D. A. Perceptual versus analytical responses to the number concept of a Weigl-type card-sorting test. *J. exp. Psychol.*, 1951, 41, 23–29.

56. GRANT, D. A., & NORRIS, E. B. Eyelid conditioning as influenced by the presence of sensitized beta-responses. *J. exp. Psychol.*, 1947, 37, 423–433.

57. GRANT, D. A., NORRIS, E. B., & BOISSARD, S. Dark adaptation and the pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1947, 37, 434–439.

58. GRANT, D. A., HAKE, H. W., & SCHNEIDER, D. E. Effects of pre-testing with the conditioned stimulus upon extinction of the conditioned eyelid response. *Amer. J. Psychol.*, 1948, 51, 243–246.

59. GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, 38, 690–696.

60. GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, 39, 35–40.

61. GRANT, D. A., HORNSETH, J. P., & HAKE, H. W. Sensitization of the beta-response as a function of the wave length of the stimulus. *J. exp. Psychol.*, 1949, 39, 195–199.

62. GRANT, D. A., JONES, O. R., & TALLANTIS, B. The relative difficulty of the number, form, and color concepts of a Weigl-type problem. *J. exp. Psychol.*, 1949, 39, 552–557.

63. GRANT, D. A., & MOTE, F. A. Effects of brief flashes of light upon the course of dark-adaptation. *J. exp. Psychol.*, 1949, 39, 610–616.

64. GRANT, D. A., RIOPELLE, A. J., & HAKE, H. W. Resistance to extinction and the pattern of reinforcement. I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, 40, 53–60.

65. GRANT, D. A., HORNSETH, J. P., & HAKE, H. W. The influence of the inter-trial interval on the Humphreys' "random reinforcement" effect during the extinction of a verbal response. *J. exp. Psychol.*, 1950, 40, 609–612.

66. GREGG, L. W., & BROGDEN, W. J. The relation between duration and reaction time difference to fixed duration and response terminated stimuli. *J. comp. physiol. Psychol.*, 1950, 43, 329–337.

67. GREGG, L. W., & BROGDEN, W. J. The relation between reaction time and the duration of the auditory stimulus. *J. comp. physiol. Psychol.*, 1950, 43, 389–395.

68. HAGGARD, E. A. On the application of analysis of variance to GSR data: I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, 39, 378–392.

69. HAGGARD, E. A. On the application of analysis of variance to GSR data: II. Some effects of the use of inappropriate measures. *J. exp. Psychol.*, 1949, 39, 861–867.

70. HELSON, H. Design of equipment and optimal human operation. *Amer. J. Psychol.*, 1949, 62, 473–497.

71. HOROWITZ, M. W. An analysis of the superiority of binocular over monocular visual acuity. *J. exp. Psychol.*, 1949, 39, 581–596.

72. HUNT, J. McV., SCHLOSBERG, H., SOLOMON, R. L., & STELLAR, E. Studies of the effects of infantile experience on adult behavior in rats. I. Effects of infantile feeding frustration on adult hoarding. *J. comp. physiol. Psychol.*, 1947, 40, 291–304.

73. JACKSON, R. W. B. Applications of the analysis of variance and covariance method to educational problems. *Dept. Educ. Res. Univer. Toronto Bull.*, 1940, No. 11.

74. JACKSON, R. W. B. Some difficulties in the application of the analysis of covariance method to educational problems. *J. educ. Psychol.*, 1941, 32, 414–422.

75. JENKINS, W. O., & POSTMAN, L. An experimental analysis of set in rote learning: retroactive inhibition of

changing set. *J. exp. Psychol.*, 1949, **39**, 69–72.

76. JOHNSON, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.

77. JOHNSON, P. O. Modern statistical science and its function in educational and psychological research. *Sci. Mon.*, 1951, **62**, 385–396.

78. JOHNSON, P. O., & NEYMAN, J. Tests of certain linear hypotheses and their application to some educational problems. *Statist. res. Mem.*, 1933, **1**, 72–93.

79. JOHNSON, P. O., & TSAO, F. Factorial design in the determination of differential limen values. *Psychometrika*, 1944, **9**, 107–144.

80. JOHNSON, P. O., & TSAO, F. Factorial design and covariance in the study of individual educational development. *Psychometrika*, 1945, **10**, 133–162.

81. JOHNSON, P. O., & FAY, L. C. The Johnson-Neyman technique, its theory and application. *Psychometrika*, 1950, **15**, 349–367.

82. KELMAN, H. C. Effects of success and failure on "suggestibility" in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267–285.

83. KEMPTHORNE, O. *The design and analysis of experiments.* New York: Wiley, 1952.

84. KENNEY, J. F., & KEEPING, E. S. *Mathematics of statistics.* Part Two. New York: Nostrand, 1951.

85. KIMBLE, G. A., & BILODEAU, E. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 150–157.

86. KOGAN, L. S. Analysis of variance-repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.

87. KUNTZ, J. E., & SLEIGHT, R. B. Legibility of numerals: the optimum ratio of height to width of stroke. *Amer. J. Psychol.*, 1950, **63**, 567–575.

88. LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.

89. LAWRENCE, D. H., & MILLER, N. E. A positive relationship between reinforcement and resistance to extinction produced by removing a source of confusion for a technique that had produced opposite results. *J. exp. Psychol.*, 1947, **37**, 494–509.

90. LEHMER, E. Inverse tables of probabilities of errors of the second kind. *Ann. math. Statist.*, 1944, **15**, 388–398.

91. LEYZOREK, M. Accuracy of visual interpolation between circular scale markers as a function of the separation between markers. *J. exp. Psychol.*, 1949, **39**, 270–271.

92. LEWIS, D. *Quantitative methods in psychology.* Iowa City: The Bookshop, 1949.

93. LIBERMAN, A. M. A comparison of transfer effects during acquisition and extinction of two instrumental responses. *J. exp. Psychol.*, 1951, **41**, 192–198.

94. LICKLIDER, J. C. R., BINDRA, D., & POLLACK, I. The intelligibility of rectangular speech-waves. *Amer. J. Psychol.*, 1948, **61**, 1–20.

95. LINDQUIST, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.

96. LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.

97. LITTMAN, R. A. Conditioned generalization of the galvanic skin reaction to tones. *J. exp. Psychol.*, 1949, **39**, 868–882.

98. LITTMAN, R. A. Latent learning in a T-maze after two degrees of training. *J. comp. physiol. Psychol.*, 1950, **43**, 135–147.

99. MANN, H. B. *Analysis and design of experiments.* New York: Dover, 1949.

100. MANN, C. W., & PASSEY, G. E. The perception of the vertical: V. Adjustment to the postural vertical as a function of the magnitude of postural tilt and duration of exposure. *J. exp. Psychol.*, 1951, **41**, 108–113.

101. McNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.

102. McNEMAR, Q. On the use of latin squares in psychology. *Psychol. Bull.*, 1951, **48**, 398–401.

103. McNEMAR, Q. Statistical theory and research design. In C. P. Stone (Ed.), *Annual review of psychology.* Stanford: Annual Reviews, 1952. Pp. 409–418.

104. MEYER, D. R., HARLOW, H. F., & SETTLAGE, P. H. A survey of delayed response performance by normal and brain damaged monkeys. *J. comp. physiol. Psychol.*, 1951, **44**, 17–25.

105. MOOD, A. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

106. MOSES, L. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.

107. NEWMAN, M., & SCHEFFLER, I. Sex differences in emotional reaction to the

news. *J. abnorm. soc. Psychol.*, 1947, 42, 476–479.

108. NORRIS, E. B., & GRANT, D. A. Eyelid conditioning as affected by verbally induced inhibitory set and counter reinforcement. *Amer. J. Psychol.*, 1948, 61, 37–49.

109. NORTH, A. J. Performance during an extended series of discrimination reversals. *J. comp. physiol. Psychol.*, 1950, 43, 461–470.

110. PORTER, P. B., STONE, C. P., & ERIKSEN, C. W. Learning ability in rats given electroconvulsive shocks in late infancy. *J. comp. physiol. Psychol.*, 1948, 41, 423–431.

111. POSTMAN, L. Choice behavior and the process of recognition. *Amer. J. Psychol.*, 1950, 63, 576–583.

112. POSTMAN, L., & JENKINS, W. O. An experimental analysis of set in rote learning: the interaction of learning instruction and retention performance. *J. exp. Psychol.*, 1948, 38, 683–689.

113. POSTMAN, L., JENKINS, W. O., & POSTMAN, D. L. An experimental comparison of active recall and recognition. *Amer. J. Psychol.*, 1948, 61, 511–519.

114. POSTMAN, L., & BRUNER, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, 39, 369–377.

115. PRENTICE, W. C. H. The relation of distance to the apparent size of figural after-effects. *Amer. J. Psychol.*, 1947, 60, 617–623.

116. PRESTON, M. G., SPIERS, A., & TRASOFF, J. On certain conditions controlling realism and irrealism of aspirations. *J. exp. Psychol.*, 1947, 37, 48–58.

117. REYNOLDS, B. Resistance to extinction as a function of the amount of reinforcement during acquisition. *J. exp. Psychol.*, 1950, 40, 46–52.

118. RYAN, T. A., COTTRELL, C. L., & BITTERMAN, M. S. Muscular tension as an index of effort: the effect of glare and other disturbances in visual work. *Amer. J. Psychol.*, 1950, 63, 317–341.

119. SATTERTHWAITE, F. E. Synthesis of variance. *Psychometrika*, 1941, 6, 309–316.

120. SATTERTHWAITE, F. E. An approximate distribution of estimates of variance components. *Biometrics*, 1946, 2, 110–114.

121. SCHOENFELD, W. N. The treatment of multiple entries in analysis of variance with three criteria of classification. *Amer. J. Psychol.*, 1944, 57, 500–508.

122. SIEGEL, P. S., & STUCKEY, H. L. The diurnal course of water and food intake in the normal mature rat. *J. comp. physiol. Psychol.*, 1947, 40, 365–370.

123. SMITH, H. F. Analysis of variance with unequal but proportionate numbers of observations in the sub-classes of a two-way classification. *Biometrics*, 1951, 7, 70–74.

124. SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Collegiate Press, 1946.

125. SOLOMON, R. The role of effort in the performance of a distance discrimination. *J. exp. Psychol.*, 1949, 39, 73–83.

126. TANG, P. C. The power function of the analysis of variance tests with tables and illustrations of their use. *Stat. res. Mem.*, 1938, 2, 126–149.

127. THOMSON, G. H. The use of the Latin square in designing educational experiments. *Brit. J. educ. Psychol.*, 1941, 11, 135–137.

128. TSAO, F. General solution of the analysis of variance and covariance in the case of unequal or disproportionate numbers of observations in the subclasses. *Psychometrika*, 1946, 11, 107–128.

129. TUKEY, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, 5, 99–114.

130. WEBB, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, 39, 1–14.

131. WILSON, J. T. The formation and retention of remote associations in rote learning. *J. exp. Psychol.*, 1949, 39, 830–838.

132. YATES, F. The analysis of replicated experiments when the field results are incomplete. *Emp. J. exp. Agric.*, 1933, 1, 129–142.

133. YOUDEN, W. J. The Fisherian revolution in methods of experimentation. *J. Amer. statist. Ass.*, 1951, 46, 47–50.

THE RESPONSE TO COLOR AND EGO FUNCTIONS[1]

ROBERT H. FORTIER
Western Reserve University[2]

It is the intent of this paper, first, to present a theory—largely derived from Schachtel (77) and Rickers-Ovsiankina (74)—and second, to examine a number of studies involving color in an effort to substantiate and clarify the theory. The reader will soon observe that the paper is heavily weighted with material from Rorschach inkblot studies. This weighting is explained by two somewhat mutually dependent factors: color, as well as a number of other variables, has been systematically treated in the Rorschach test, and there have been a multitude of papers written in which the Rorschach test was featured. Whenever possible, parallel studies which use other techniques employing color will be introduced.

DERIVATION OF A THEORY
Schachtel

Schachtel feels that the experience of color and the experience of affect have two important characteristics in common: ". . . the passivity of the subject, and the immediacy of the relation object-subject" (77, p. 399).

Two examples from Schachtel may clarify this analogy.

1. An individual enters a room in which there are two designs. On one wall is a large blob of color. On the opposite wall is a large design in black and white. The blob of color is immediately perceived, almost without conscious attention. The individual is aware only of color. The design in black and white requires directed attention before it can be perceived.

2. An individual becomes angered. He strikes out blindly at his antagonist, without regard for the consequences of his act. He is aware only of his anger and an object upon which to vent this anger.

The two examples are extremes. One may imagine more moderate behaviors in the two situations described above. An individual entering the room immediately perceives the blob of color, but his perception also encompasses the contour, and some analysis of its shape occurs. An individual becomes extremely angry but acts upon the situation in such a manner that the tension produced by the anger is reduced without violence being done either to the stimulus of the anger or to the individual himself.

It is a primary task of the ego to control and direct affective reactions, whether produced by drives originating from without the ego or from within it (40). The individuals in the two situations described by Schachtel were completely passive. They were literally swept away by their experiences. Their egos, which should have channeled and controlled the aroused experiences, failed in their tasks. The behaviors of the individuals in the situations described by the writer may have retained some elements of

passivity, but this passivity was considerably moderated by ego control. It is evident that passivity does not refer to the overt behavior of the individual, but only to the relation between his affective drives and his ego. Schachtel points out that the affective experience is a conscious one, regardless of the passivity of the ego. Where there are no affects, there is no consciousness of drives.

Rickers-Ovsiankina

Rickers-Ovsiankina, after reviewing the literature (74), concludes that an individual's response to color can give considerable insight into the degree of permeability of his ego. That individual whose ego is responsive to the outside world will respond to color. As the degree of permeability increases—that is, as the boundary between the ego and the outside world lessens in strength—the individual will respond more to color per se.

Both Schachtel and Rickers-Ovsiankina stress the fact that the extratensive individual is responding to outside stimuli and that there is a lessening in the spontaneity of the individual. However, one misses an important point if he concentrates upon this particular elaboration in Rickers-Ovsiankina's paper. Schachtel feels that the individual who responds to color per se not only possesses an ego which readily responds to the outside environment but which also is less capable of exerting control upon affective drives having internal origin. He feels then that the permeability of the ego is a two-way affair, for there is also a more direct release of affective drive upon the external environment.

The Concept of the Egocentric Individual

Let us diverge for a moment to consider another aspect of the individual who responds to color per se. Such a divergence will serve the dual purpose of clarifying the points presented thus far, and of facilitating interpretation of certain of the studies which will be presented later.

The individual who responds to color per se has been called egocentric by a number of authors (11, 16, 41, 42, 54, 68). In the light of the preceding discussion, just what does "egocentricity" mean? Warren (88) defines the term "egocentric" as follows: "disposed to dwell on oneself and to view every situation from a personal angle" (88, p. 89). As a synonym, he gives the term "self-centered." Do Schachtel and Rickers-Ovsiankina actually consider the individual who responds to relatively undifferentiated color "egocentric"? Under Schachtel's scheme the individual adopting this mode of color response may behave in one of two ways, or both.

The individual will adapt to his environment, behaving as it dictates. If an environmental configuration directs action in one way, he will act in that way. If action is directed another way, then the individual will again modify his action to conform to environmental pressure. This individual clearly cannot be called "self-centered," for he is most likely to perform in the manner in which others wish him to perform. But Schachtel further feels that the individual responding to relatively undifferentiated color may also release his affective drives in a relatively undifferentiated manner upon the environment without regard for that environment. It is this latter behavior which might best be called "egocentric." Nevertheless, Schachtel makes it clear that an individual who behaves in this manner is not channeling his affective drives. There

is a more or less direct interchange between his affective drives and the environment. The individual's behavior, therefore, rather than being egocentric, is relatively removed from the control of the ego. The ego plays a relatively unimportant role as to the object upon which the affective charge is released and the manner of its release. It is certainly far from a deliberate (that is, ego in origin) attempt to ignore the feelings of others that results in the sometimes antisocial and inconsiderate behavior of this individual. It is the result of a fundamental incapacity of the ego to direct and control the affective charge in a realistic manner.

STATEMENT OF A THEORY

The experience of affect and the experience of color are quite comparable. Thus one may examine the less obvious of the two, affect, by the response to the more obvious, color. The experience of affect is passive. The degree of passivity is determined by the degree of control exerted upon the affective charge by the ego. That individual who responds to relatively undifferentiated color possesses an ego which is less able to control and channel affective charge. Such an individual lacks spontaneity of action and readily adopts the color of his environment. It is a logical corollary that an individual responding to relatively undifferentiated color may release affective charge upon the environment in a more or less undifferentiated manner. Further, his perception of affective charge in others may also be relatively undifferentiated. That is, the individual may sense very acutely the presence of affect in others without being able to differentiate it, to identify its nature. Thus, when someone becomes angry with him, he may only be aware of the existence of a powerful

affective state in his antagonist. He may not know what is the nature of this state. However, the particular course of action followed by the individual depends upon his total personality configuration.[3]

THE STUDIES

A theory, no matter what its logical integrity, must be tested by data set forth in a variety of studies. Rickers-Ovsiankina in her paper (74) to which reference is made above, reviews a number of articles and the present writer does not intend to duplicate her bibliography to any great extent. The studies included here vary from those dealing with normal individuals and their development through those concerned with individuals suffering from organic brain damage.

Since a number of studies to which reference will be made later concern the Rorschach test, brief mention will be made of the treatment of color by users of the test. Those desiring a comprehensive exposition of the treatment of color by Rorschach investigators may refer to any of the standard texts (11, 12, 16, 54, etc.) or to the normative studies of Hertz (41, 42). Suffice it to say that the scoring of a response to color depends upon the degree of structure imparted to

[3] The reader who refers to the original manuscripts by Schachtel and Rickers-Ovsiankina will see that for the most part the statement presented here is simply a more succinct and perhaps clearer presentation of some of the ideas formulated in the two papers. The suggestion that a blunting of the perception of affective drives in others accompanies the response to relatively undifferentiated color may be considered the most important addition. The writer feels that at this stage of knowledge, little may be gained from further additions to the theory, but that the theory and the manner in which it is used in the present paper can serve to promote more rigorous investigations, in that way contributing to an advancement of the theory.

the color or the degree of integration of color with form. An undifferentiated or unstructured color response is one determined solely by the color of the blot. This is scored as *C*. A response which is determined principally by the color but which is structured to some extent is scored as *CF*. A response in which the color is quite integrated with form is scored as *FC*. The nature of the affective experience which is represented by each of these scoring categories is apparent from what has been said above. The three color factors have been assigned numerical weights as follows: *FC*, .05; *CF*, 1.0; *C*, 1.5. These weights were originally suggested by Rorschach, and have been used extensively. Rorschach's belief was that, since *C* represented a more powerful and uncontrolled affective drive, it should have the greatest weighting and *FC* the least. Weighting finds applicability particularly in determining the so called stability or control ratio, calculated by the formula

$$FC - (CF + C).$$

The Normal Picture

The criteria for normalcy may be statistical, may be based upon psychological theory and knowledge of dynamics, or may be philosophical. It is not the purpose of this paper to delve deeply into the dynamics or total configuration of any group. Therefore, for the criteria of normalcy (in relation to the affective aspect), the following should suffice.

First, it seems logical that in our society an individual must have affective drives and affective relationships with others, and must be relatively in control of these drives and relationships. That is, the direction and manner of release of affective charge must be ego-controlled. He must also be able to interpret and in-

tegrate the affective behavior of others. Second, the world must not be so firmly fixed and structured in his ego that he can not be moved or partially influenced by the particular configuration of the environment. To a certain extent, his affective behavior towards others should not continually require meditation and deliberation. Third, affective reactions directed towards the environment by the individual cannot be overly gross and undifferentiated, nor can his perception of and reaction to the affective behavior of others be undifferentiated or gross. The type of color response on the Rorschach Test representing each of the above delineations is obvious: the first by *FC*; the second by *CF*; the third by *C*.

The Normal Adult

Rorschach findings. What is found by Rorschach examination of normal individuals? There are several sources of information, among them the standard texts and normative studies by Hertz already cited. Klopfer and Kelley (54) suggest that the normal adult should give some color responses, but that the sum weight of *C* and *CF* responses should not be higher than the sum weight of *FC* responses. They feel that a crude *C* response, one which is not descriptive or symbolic, is a pathological sign. (This view is not shared by Hertz.) Beck (11, 12) gives no specific norms for any group, but presents the psychological significance of the various types of color responses and suggests some individuals who would give them. The sign of the healthy individual is *FC*. Such an individual is mature and can establish affective relations with other individuals. This individual may give a number of, or a few *CF* responses, but the sum weight of the *FC* responses should approximately equal or exceed

the $CF+C$ sum weight. C is suggestive of regression.

In a major normative study by Hertz and Baker (42) it was found that an average of 3.7 color responses is given by 15-year-old boys and girls. No averages are given for CF or C responses. However, the range for CF responses is 0–2; the range for C responses is 0–1. Thus, one C response would not be considered pathological. The average sum weight of color responses is 2.8. Balance in favor of FC is indicated by the weighted ratio $FC-(CF+C)$, which is $+0.54$. The implication is clear that Hertz and Baker would expect normal adults to have at least as balanced a ratio in favor of FC, or a higher one. Steinzor (84), in testing a presumably normal college group, found a balance in favor of FC. In a healthy individual, a sign of integration is a balance in favor of FC, with a total color weight of 3.0.

Beck, *et al.* report a normative study based on a fairly large number of adults. Concerning the color factors, they conclude:

Of most interest is the weighting in the direction of CF; and FC, in that order; and the comparatively small instance of undiluted C. The population of which this sample is representative may, therefore, in respect to affectivity, be described as having made some progress towards maturity and towards capacity for social rapport. Yet they are slightly more labile than fully stabilized. On the other hand, the quantity of infantile egocentricity is relatively small . . . unstable, easily excited, but resisting undisciplined violence (13, p. 259).

The reader may wonder whether the theory has after all been formulated incorrectly, with this sudden finding of a predominance of CF over FC. However, there is another explanation which, to the writer, seems rather feasible. It may be, as Beck

suggests later in the study, that the dynamic configuration of the individual and the society has changed. The color factors reflect this change. A clarification may result if one attributes to these findings the psychological significance formulated by the theory stated in the present paper, without allusions to maturity or infantile reactions. The presence of FC indicates a capacity for ego-controlled affectivity and the capacity to integrate and interpret the affective behavior of others. The excess of CF over FC indicates that the individual may have an immediate reaction to the environment and may be considerably influenced by it. The very small instance of C is a countersign against gross and undifferentiated affective behavior and perception. In other words, although the present adult may react and respond far more readily to his environment than he once did—if the earlier diagnoses were correct—he still has the capacity for constructive affective behavior.

Studies involving painting. The majority of studies involving easel and fingerpainting had children as subjects. However, Waehner (87), from his work with college students, developed several indices in regard to emotional balance, control, compulsion, constriction, etc. In constructing these indices, Waehner deliberately paralleled to a considerable extent certain Rorschach procedures. Superior emotional balance is reflected by a color variety of three to six, and a relationship of color to form of $5C:4F$. Constriction is indicated by a small variety of or no color.

Napoli (65, 66) will not be discussed here since he is more concerned with the particular hues rather than color in general. Those interested in a broad survey of paint-

ing are referred to a recent article by Precker (72).

The Mosaic Test. This instrument has been gaining in interest among individuals with varied approaches to study of personality, but as yet few articles have been published. Wertham and Golden (92), while more concerned with the forms of the designs reproduced, expect normal individuals to produce designs harmonious in color and distinct in configuration. Diamond and Schmale (26) say that normals may have a very wide range in the use of color from primitive and crude designs in color to extremely artistic use of color. In comparing data obtained from the Mosaic Test with that obtained from the Rorschach test, the authors conclude that there is a tremendous discrepancy between the results obtained by the two instruments. This discrepancy and some possible theory underlying the varying performances obtained with the two instruments will be elaborated when psychotic modes of adjustment are discussed later in this paper. Lowenfield (60) confirms Diamond and Schmale's (26) findings on normals.

A Genetic Approach

Many investigators feel that a correlate of increasing chronological age is increasing emotional control. From an examination of genetic studies, therefore, one should be able to determine, first, how the individual performs at different ages, and second, what is the psychological significance of this performance. Jersild reviews a number of studies of emotional development and concludes: "The data now available from direct observation or experimental study do not provide the basis for a systematic account of normal and immature emotional behavior at various age

levels" (45, p. 760). Therefore, what material can be cited here is admittedly sketchy and incomplete.

Pratt, Nelson, and Sun (in 45), as a result of their studies of neonates and slightly older children, stress the point that "generalized reactions predominate over specific reactions in early childhood and the fact that distinctive patterns are difficult to detect." Taylor (in 45) also stresses the undifferentiated nature of emotional reactions in a study of children aged *one to twelve days*. Sherman (in 45) arrives at a similar conclusion and expresses the belief that "with the passage of time the child's behavior becomes increasingly differentiated and adaptive." Jersild (45) points out that the manifest emotional behavior may change through the changing nature of the emotional problems with which the child is confronted.

Gesell (32) comments briefly on the problem of emotional development. We may compare what he says with what may be inferred from other sources.

The three year old attempts to conform and please, "as though he were sensitive to the demands of the culture." Suggestions are accepted more readily. He may prefer the companionship of other children but as yet is incapable of verbalizing his desires. He can play with children for a while, but may suddenly attack them. He is at least somewhat susceptible to social suggestion. He studies the facial expressions of individuals in his environment and attempts to interpret them. "He is capable of sympathy." Similar sketches are drawn for the four- and five-year-old. The progress made by the three-year-old child is continued, but the pattern may change slightly. "Three has a conforming mind. Four has a lively mind. Three is assentive; four assertive." Five shows much more definiteness, concreteness. Gesell calls this age a plateau.

The Rorschach findings. Unfortunately, the sources of information are

again meager. Klopfer and Marguiles (55) made a study of children aged two through six years. They present percentages of children in each age group who use the various color scoring categories, and the average number of such responses given by each child. The results on the color factors are reproduced in Table 1. The authors' findings that at the age of six *FC* dominated the types of color responses should be particularly noted.

A second group of children, aged three through seven years, was studied several years later by Ford (30). To facilitate interpretation and comparison with the results obtained by Klopfer and Marguiles, a summary of her results on the color factors is presented in Table 1.

TABLE 1

Percentage of Children of Different Age Levels Giving Color Responses and the Average Number of Color Responses Given

Ages	FC	CF	C*	FC	CF	C	Cn†
			(*After Klopfer*)				
2	12	12	29	.29	.23	.59	.23
3	37	24	46	.87	.41	1.04	.48
4	27	45	30	.88	1.24	.54	.09
5	54	67	41	1.29	1.44	.75	—
6	65	43	30	2.08	.78	.69	—
			(*After Ford*)				
3	28	32	12	.5	.5	.2	.7
4	44	56	15	.8	.8	.2	.7
5	56	36	36	.8	.5	.8	.4
6	57	61	35	.7	1.0	.5	0
7	52	57	43	1.1	1.2	.7	0

* In terms of per cent.
† In terms of average number.

A principal source of discrepancy, at least in reporting the percentages of children giving the responses, is derived from the fact that Ford very

carefully distinguished between pure *C* responses and color naming (*Cn*), that is, where an individual simply gave the name of the color. Klopfer and Marguiles made no such distinction in reporting percentages.

An explanation of the diverging results obtained by Ford and Klopfer and Marguiles is difficult to find. Both studies were based upon children who could be expected to have comparable socioeconomic backgrounds. One source of variance may lie in the particular scoring bias of the different investigators. The records used by Klopfer and Marguiles were submitted by a number of different investigators, however, and several other investigators confirmed the findings by Ford. It may be that a difference in time—the year in which the investigations were made—is a contributing factor. The study by Klopfer and Marguiles was reported in 1941; the study by Ford in 1946. Obviously both investigations took much time to prepare. If it can be determined that the study by Klopfer and Marguiles antedated by any great period of time that by Ford, this difference, when interpreted in the light of the recent study by Beck (13), may point to a fairly rapid altering in the dynamic configuration of the individual and the society. This is particularly true of the discrepancy obtained in regard to the dominance of *FC*.[4]

Studies involving painting. The text by Alschuler and Hattwick (3) undoubtedly represents the most com-

[4] The mounting percentages of children giving color responses associated with increasing age are readily explained if one recalls Schachtel's (77) injunction that affects are the conscious representations of drives. Obviously, a child becomes more aware of his drives with increasing age. Therefore, a very young child giving many color responses is showing signs of "precocity" rather than "infantilism."

plete and recent major work concerned with painting and children. The authors' findings tend to confirm Rorschach test results with children. Children of three or four are quite interested in color and tend to use it without great regard for form. With increasing age, the color tends to become more and more integrated with form. In studying groups of children, one of which uses a great deal of color and one of which is more concerned with form, the authors found that children concerned with form were more self-controlled, more concerned with external stimuli, and had a higher frequency of reasoned (in contrast to impulsive) behavior than those using much color.

Epstein and Schwartz (29) found that the number of colors used reflects the emotional development of the child; those using under four colors having poor emotional development, lack of drive, and perhaps constriction. Overcontrol or retarded development is indicated by a predominance of form over color. That interest in and use of color declines after the child reaches a certain age was confirmed by Blum and Dragowitz (15).

Thus it is seen that although an integration of form with color in painting—at least to a certain extent—may be expected from children who have acquired some facility with emotional control, there is a certain time lag involved as to when such integration occurs. A suggestion is that the two methods of studying emotional development—the Rorschach test and painting—may measure different aspects of this development. Since the writer feels that, to a considerable extent, the Mosaic Test and easel painting are functionally comparable, a discussion of the discrepancies obtained through use of

the Mosaic Test and the Rorschach test will be discussed later as already indicated.

Contribution of the Genetic Approach and the Adult Picture

The task is now to re-examine the psychological significance of the color factors in the light of the evidence presented in these two sections. Beck (12), in elaborating the psychological significance of the pure C response, says: "This is the reaction mode of the infant, who does what he pleases—screams, demands food, kicks, voids without regard to time and place. Response to feelings is exclusive and instant" (p. 30). Jersild (45) holds such reactions as described by Beck as typical of very young children, neonates, and children up until perhaps the age of two. The studies reviewed by Jersild (45) definitely point out that emotional control rapidly increases with increasing age. A neonate obviously cannot be given a Rorschach test. If Ford's results (and those of investigators reporting similar findings) may be accepted as at least typical of a certain class of children, it is seen that three-year-old children as a group gave but 0.2 C responses. It then becomes difficult to see upon what basis an interpretation such as Beck's is founded. Gesell (32) paints a picture of the three-year-old as an individual who has gained considerably in emotional control, but who is still markedly dependent upon the desires and wishes of those in his environment. He is also prone *upon occasions* to attack quickly and violently those individuals around him. He is interested in studying and trying to interpret the facial expressions of those around him.

From this picture, something comparable to Ford's results could have

been expected. Of the three-year-old children, 28 per cent gave *FC* responses. *CF* responses predominate, while *C* responses are last in both per cent and average number. The finding that *CF* predominates at this age level, coupled with the statements that the three-year-old child is "assentive" and is ". . . sensitive to the demands of the culture . . ." (32, p. 36) corroborates the interpretation of the *CF* response presented by the present writer. The statement by Gesell (32) that the five-year level constitutes a plateau coincides quite well with the finding by Ford that it is only at this age level that a predominance of *FC* over *CF* is found. The significance of the *C* response also seems to mean what it was sketched as meaning in the introductory theory—affective charge not controlled or directed by the ego.

The significance of the *Cn* (color naming) response is still to be determined. Anyone observing the relations of young children and parents will find, during a certain phase in the child's development, parents pointing to different colored objects, regardless of their shapes, saying, "this is green, this is yellow," etc. It therefore seems logical that, when a young child is presented with a new type of game, he will point out certain areas on the card and say, "this is green, this is yellow," etc. Or, if one assumes that the hypothesized relation between affect and color holds also in this situation, another interpretation is possible. Gesell (32) points out that the child studies the facial expression of those around him and tries to interpret them. The interpretation may be of the nature of nosology: that is, the attempt may be to identify, without the ability to act upon the interpretation. Then, may not the child, after looking at his father, say to himself, "he is angry," or "he is sad," without having the ability to act constructively upon his identification?

No contradiction is found between the psychological significance that is attributed to the color factors when used by the child and the psychological significance that is attributed to the color factors when used by the adult. The writer feels that if one is successful in defining the psychological significance of a variable, and this significance remains constant whether the variable is used by an adult or a child, it is better to use this definition than to attempt to define the variable in terms of one or the other chronological referents.

Nonpsychopathological Deviations

The Institutionalized Child

Turning from normal individuals raised and living in a normal environment, attention can be focused upon those individuals—more specifically, children—who have spent the greater portion of their lives in institutions. As a result of a study by Goldfarb (34), it was found that more such children give the pure *C* response, and that more of them exhibit the unbalanced ratio of sum weight *CF* + *C* greater than *FC* weight. Goldfarb suggests that this is an indication of the lessening of rational control and a greater emotional immaturity.

Goldfarb and Klopfer continue this analysis, concluding that:

The institutional group thus shows deficiencies in rational control, in more abstract forms of thinking, in drive for intellectual and social attainments, and in emotional maturity. In a group with such psychological tendencies one would, of course, expect problems involving restlessness, inability to concentrate, and poor adjustment. In addition, all of the

above listed Rorschach trends among the "institution" children are associated with an *air of passivity* (italics added). In other words, the children of this group give little of themselves though superficially they are adjusting to reality requirements (36, p. 93).

The writer feels it significant that Goldfarb and Klopfer find an air of passivity in institutionalized children, for were one to attribute the usual significance to an unbalanced $CF+C$ ratio, a far from passive attitude would be expected—vigorously negativistic, impulsive, willful, etc. What this ratio actually seems to suggest here is a greater inclination to be moved and swayed by the environmental configuration in which the children find themselves. It suggests a more undifferentiated emotional approach to the environment, possibly a rather diffuse emotional reaction towards everyone with whom the children come into contact. Such behavior would logically follow from the nature of the institutions in which the children find themselves forced to live. Their behavior is in virtually every respect governed by more or less impersonal rules and regulations. Their relations with adults are limited simply because there are so few adults in the institutions that personal contact is quite difficult. The suggestion follows naturally that for one to have sufficient intellectual control of affectivity one must have the opportunity to learn and develop this capacity. It is probable, then, that the nature of one's affective life is dependent, above native endowment, perhaps, upon the ability and opportunity to learn.

The Delinquent

In regard to this last postulate—affective control and opportunity for learning—the findings of those investigators concerned with juvenile delinquents may have some bearing. "Burt holds that marked emotionality is the most frequent and most influential of all the psychological characteristics of the delinquent" (27, p. 129). This statement sums up succinctly an attitude towards the genesis of delinquency which prevailed for a considerable period of time, and perhaps prevails in certain quarters now, judging from the studies still concerned with the relation between emotionality and delinquency.

Rorschach findings. The Rorschach findings may prove rather startling to those investigators still adhering to the classical theory of delinquency quoted above. Endacott (28) in a study of 100 delinquent boys—average age, 14 years—found a restriction of color, lower FC, and lower CF when his results were compared to those of other investigators. The normative study by Hertz and Baker (42) suggests that a *sum C* weight of one to one and one-half points higher could be expected from boys of this age. Boynton and Walsworth (18) report a study of 47 delinquent vocational school (reform school) girls of approximately high school age. The authors compared the results obtained from the Rorschach protocols of the delinquent girls with those obtained from girls attending a high school located in a favorable section of the same town. In regard to color, the delinquent girls scored lower than the high school girls in all respects. The so-called impulsivity ratio was more in favor of $CF+C$ in the high school group than in the delinquent group. A number of earlier studies reporting excessive emotionality in delinquents were reviewed by Schmidl (79) and criticized because of inadequate sampling and other factors. He points out the suggestion by

Beck that delinquents can be either extratensive or introversive.

These later Rorschach findings suggest that some cause other than marked emotionality must be postulated to explain the antisocial behavior of delinquents. There is a suggestion that the inability to establish rapport (or lack of ego-controlled affective charges) cannot be accepted as a general factor in explaining delinquency. Boynton and Walsworth (18) feel that one should be quite careful in using personality aberrations as explanations for delinquent behavior. Endacott sums up his findings by saying, "these are marks of a rigid, stiff-geared sort of personality that has been created to withstand strong pressures and frustrations" (28).

The implication for affect or color theory is clear. If it can be shown that a group having a somewhat lower use of color and a more stable color ratio than "normal" individuals indulges in strong, "self-centered," antisocial behavior, it is a logical deduction that it is not the affective relationship with the environment which should be postulated as a causative factor. There is a further indication that the presence of ego-controlled affectivity or "capacity for rapport" suggests nothing more concerning the individual than that he can direct and control his affective charges and can interpret and integrate the affective behavior of others. "Capacity for rapport" suggests nothing concerning the content of the behavior of the individual. The delinquent, it would appear, is acting in accordance with his picture of the reality, i.e., in accordance with his ego. A logical corollary of the above deduction is this: where antisocial, self-centered behavior appears concomitantly with extratensive, unbalanced use of color,

one should look beyond affect for an explanation of this behavior.

A study using fingerpainting. It has often happened that where one method failed to provide insight into a particular problem, another method has succeeded at least partially. The method of approach to and the products of fingerpaintings of delinquent and high school youths were compared by Phillips and Stromberg (69). While a number of significant differences are reported in their study, only one need be discussed here. Thirty-six per cent of the high school group used only one color on the first performance. Sixty-four per cent of the delinquents used only one color on the first performance. This difference is not quite significant at the .05 level. However, on the second performance, 4 per cent of the high school students used only one color, while 60 per cent of the delinquents continued to use only one color. The result of this comparison is highly significant statistically.

If it can be assumed that one's handling of color is indicative of his affective life, the nondelinquent, and the delinquent to a greater degree, might here be showing a certain amount of shock when confronted with a new—and perhaps affective—situation, and thus respond in a somewhat stereotyped manner. However, the nondelinquent shows a considerable degree of recoverability and a capacity for a wide variety of response. The delinquent, on the other hand, continues to show a stereotyped reaction. When one recalls a deduction made from the evidence reviewed concerning institutionalized children, it is possible that the environment in which delinquents live does not make it possible for them to learn a variety of emotional reactions. Thus, while it may not be any fundamental lack of capacity for emotional

rapport or ego control of affect which contributes to their antisocial behavior, it may be an inability of the delinquents to vary their emotional response. This suggestion accords with the reduced use of color in general on the Rorschach test and the reduced use of *CF* by delinquents.

Psychopathological Deviations

Investigators have found that considerable insight may be obtained into certain dynamic relationships and functions by studying individuals exhibiting more or less psychopathological reactions. It is reasonable to expect that a comparable result may be obtained here by reviewing the behavior of such individuals towards color.

Alcoholics

Using the Rorschach test, Billig and Sullivan (14) found that the affective picture presented in the use of color by alcoholics is of considerable prognostic value. In reference to those alcoholics who over a period of time showed the least favorable prognosis, the authors conclude: "In 80% of the cases factors indicating impulsive emotional behavior are stronger than those expressing smooth adjustment to environmental influences" (14, p. 124). However, if their table is accurate, the impulsive use of color is reflected primarily by the use of *CF* rather than *C*. There is a marked reduction in the appearance of *FC*, and consequently the color ratio is unbalanced in favor of *CF*. The authors feel that their results confirm a previous study by Bowman and Jellinek, who made the statement that "the chronic alcoholic shows a comparatively weak restraint, poor mental poise and stability, difficulties in controlling his mood swings and desires, combined with a lack of attention" (14, p. 124).

It is immediately apparent that the statement cited from the paper by Bowman and Jellinek presents an interpretation which is far more comparable to that which would be made by the present writer (drawing upon the theoretical outline sketched) on the basis of the evidence presented by Billig and Sullivan than the one actually made by the latter authors. The comparative lack of *FC* responses among the alchoholics with a poor prognosis, coupled with the overabundance of *CF* responses, suggest that these alcoholics are particularly affected by the environmental configuration in which they are immersed and are lacking the capacity to integrate and control the affective drives which are therefore readily aroused in them. The appearance of pure *C* in these cases would simply add a more unfavorable touch by suggesting relatively undifferentiated and diffuse emotional reactions and interpretations.

Enuretics

Enuretic children under ten years of age were shown by Goldfarb to have a high *sum C* total "with a conspicuous excess of *CF* and uncontrolled *C* responses. Emotional development at a primitive, infantile, impulsive level is suggested" (33, p. 30). Specific figures are not given. However, a glance at the studies by Ford (30) and Swift (86) show that, when color naming is excluded, pure *C* is not a particularly frequent type of response. To be sure, *CF* outweighs *FC*, but some *FC* does appear. A high frequency of *C* and *CF* responses does not seem to be typical of children of the ages studied so far, and considerable difficulty would be encountered in the task of determining whether such a reaction is typical of infants in the technical sense of the term, i.e., from birth to two years.

For one to call this excessive use of unbalanced color *infantile* is to use an analogy which is dubious and which may never be confirmed. Even were the analogy correct, to say that something is infantile is not in itself particularly expressive because little concerning the dynamics of the function involved is suggested. What this mode of usage of color by enuretics suggests (as has been shown in other cases) is a considerable degree of influence by the environment, with a relatively diffuse reaction towards it, and a reduction in the ego control of affect. Study of the ego of the enuretic might be more revealing of the dynamics of this particular dysfunction, i.e., enuresis, than analysis of the affective factors alone. Then, one may infer that this undifferentiated type of reaction and excessive influence of the environment contribute to the development of the symptom. Goldfarb's study does contribute to our knowledge of the ego content of the enuretic. Only the interpretation of the color factors is in question.

The Hysteric

Schafer (78) suggests that a characteristic of the hysteric is a predominance of $CF+C$ over FC in the Rorschach test. A further characterization is a "minimization of active and independent ideation as a means of coping with problems" (78, p. 33). Such characterizations tend to substantiate the theory that a less integrated color response suggests a greater susceptibility to environmental influence and a lessening control of one's affect.

The problem of egocentricity enters the picture. The present writer feels, as he earlier proposed, that the presence of relatively uncontrolled color is neither an indication of nor a countersign against egocentricity as it appears in hysterics and others. He feels that it would be much more feasible to regard the behavior of the hysteric which leads others to call him egocentric as a quality of the ego content, or picture of reality, of the hysteric rather than of his affective life.

Schizophrenic Adjustment and the Rorschach Test

Among the psychotic modes of adjustment, that of schizophrenia has attracted the most attention. No detailed review of the dynamic picture of the schizophrenic will be attempted since the task would be complicated by the belief held by many investigators that schizophrenia is not a single disease entity. Beck (9, 11, 12), Kelley and Klopfer (49), Klopfer and Kelley (54), Rickers-Ovsiankina (73), Kisker (52), Stern and Malloy (85), Kendig (51), and others, feel that a characteristic of schizophrenia is an overwhelming imbalance in color responses in the direction of $CF+C$. The total color weight may be large or small. As Beck (9) suggests, affect is not absent or even negligible in the schizophrenic as was once felt to be the case.

The color factors indicate first, according to the present writer's introductory scheme, that the schizophrenic is considerably influenced by the environmental configuration in which he finds himself; second, that his perception of affective life in others and his affective reaction to the environment are undifferentiated and gross. The inappropriate affective reaction frequently noted in the schizophrenic is explained by the finding of pure C in the record. If the schizophrenic cannot interpret correctly the nature of the affective situation with which he is confronted, this, coupled with the fact that he has

little ego control of his affectivity, certainly would make it rather a coincidence if appropriate emotional reactions did result.

To postulate that the schizophrenic is considerably influenced by the environmental configuration may deviate somewhat from the typical concept of the schizophrenic as separated from reality. The mere fact that the schizophrenic is considerably influenced by the environment does not indicate that his reactions to the environment will be realistic. The content of his behavior is determined by the content of his ego. It is agreed that the ego of the schizophrenic contains far from a realistic conception of the environment. Therefore, this environmental influence operating upon a distorted picture of reality would only add to the confusion and bizarre reactions of the schizophrenic. The susceptibility to environmental influence might explain the finding by several clinicians that a great deal of what occurs around the schizophrenic in a catatonic stupor is frequently remembered by the schizophrenic when he recovers from the stupor.

Schizophrenic Adjustment and the Mosaic Test: The Suggestion of a Theory

Diamond and Schmale's (26) investigation of schizophrenia by use of the Mosaic Test, concerned with the dynamics of the schizophrenic, theory of the Mosaic Test, and affect-color theory, may prove to be of considerable value. The authors found that the schizophrenic completely disregarded the color of the pieces in the construction of his design.

The color defects of the schizophrenic deserve special discussion. Color rejection or color disregard appears very early in this disease even though the personality and the Mosaic pattern are seemingly well-integrated. . . . The Mosaic pattern is exactly as if it were constructed by a totally color blind individual. It might be called a psychological color blindness. . . . It was very difficult to compare the color responses to the Rorschach and the Mosaic Tests in individual cases, and little consistency between the two were shown (p. 246).

An obvious source of this inconsistency with the two tests is in their basic dynamics. The Rorschach inkblots are often called unstructured, ambiguous, undefined. This is true, but only in a certain sense. They do exist; they do have very definite and unchanging configurations. No matter how the subject looks at the blots, there is no change in the actual shape of the blots. The Mosaic Test is completely different. Before the subject are a large number of little blocks of many different shapes and colors. He moves them around and can put them back together in a multitude of different ways. (It can be seen that the functioning of an individual in regard to easel- or fingerpainting is comparable to what it is with the Mosaic Test. The writer feels that the task presented by the Mosaic Test is more difficult than that presented by finger- or other painting.)

Many investigators feel that the task demanded of the individual confronted with the Rorschach test is essentially a creative one. The same can be said of the Mosaic Test. But is not the nature of the creative activity, the basic mechanism of the creative activity, extremely different from one test to the other? The creativity involved in the Rorschach test is exclusively an associational one. Upon being presented with an unchanging configuration, the individual is asked to call upon the content of his ego for a concept which corresponds more or less to the actual shape of the configuration. The individual is not asked to alter reality.

He changes nothing in the external environment nor does he create anything in it. The task involved in the Mosaic Test is also associative, but it is far more than that. Before one constructs a definite thing in the external environment, he has a more or less defined image of that thing in his mind. The nature and variety of what is brought forth depends, as it does in the Rorschach test, upon the content of the ego and its associative facility. But the development of the concept or image must be followed by an alteration in the external environment. The resulting product depends upon manipulative skill to some extent, but to a greater extent upon the individual's capacity to translate his more or less well-defined image into concrete reality.

How does this difference in the basic mechanisms involved in the two tests affect the product? The response to color only is of pertinence here. Consider the schizophrenic individual whose capacity for ego-controlled and ego-oriented affective charges is considerably reduced, and whose emotional responses are becoming increasingly gross and undifferentiated. On a piece of cardboard before him, he sees a blob of color. Not much capacity or skill is required for this individual to call forth a vague, structureless association. More often than not, if he succeeds in developing a structured response, the response will not fit the configuration confronting him.

In the Mosaic Test, for the individual to handle color properly— that is, if he is not to ignore it—he must first be able to visualize or conceive an image or configuration involving color and then reproduce this image by an alteration of the external environment. An individual whose capacity for ego-controlled affectivity is greatly reduced would

have much difficulty in conceiving an affective situation and more in manipulating it in the external environment. The product of this individual would, of course, show not only a disregard of the color of the Mosaic blocks, but very poor form as well.

But an individual who is not quite sure of himself, who is becoming aware that something about his handling of affectively charged situations is not quite right, would, even if he could conceive a fairly adequate affectively charged image, hesitate to bring this image forth, to reproduce it in the external reality where it would be visible not only to himself but to others. Since Diamond and Schmale (26) found that even in the very early stages of schizophrenia color was ignored while the capacity to integrate form was relatively intact, the suggestion is strong that one of the very first signs of the schizophrenic process is an uncertainty, perhaps even consciously realized, that one is losing his capacity to handle affective situations.

The writer thinks it will be agreed that capacity to handle emotionally toned situations will vary among individuals who do possess facility to conceive and interpret such situations. If this is so, and if the writer's analysis of the process underlying the Mosaic Test is correct, it is not difficult to see why results obtained in one test are not comparable to those obtained in the other. Thus there are two aspects to affective situations: the ability to conceive and interpret them; and the ability to handle them in the external environment. The reasoning followed here tends to be confirmed by a study of institutionalized children conducted by Colm (24). These children also used color indiscriminately. It is obvious that institutionalized children have limited scope and opportunity

for learning to handle affective situations.

The Feebleminded Individual

Davidson and Klopfer (25), Kelley (48), Abel (1), and Werner (90) agree that on the Rorschach test the mental defective, although he may use less color, uses it in an unbalanced fashion, i.e., $CF+C$ greater than FC. Abel (1), in studying defectives showing the least inclination to succeed in school, found them to give more such unbalanced records than successful defectives. He concluded that the former are more susceptible to "stimulation from the external environment without adequate control of the situation."

Studies of Individuals Having Cerebral Disorders

Epileptics. Guirdham (38), Arluck (6), and Kelley (50) agree that the epileptic gives fewer color responses on the Rorschach test than the normal individual. However, these responses are definitely unbalanced in favor of $CF+C$. The suggestion is that the epileptic has little emotional communication with the environment, and that such communication as does occur is not under ego control, tending to be very gross and undifferentiated. It is of considerable significance that Drohocki (cited in 70) found, upon repeated examination of epileptics beginning immediately after seizure, an extremely dilated picture, with evidently a number of color responses, the majority of which were unbalanced in favor of $CF+C$. Stainbrook (82) found somewhat similar occurrences however, with first the appearance of Cn, then CF and FC.

The findings of Drohocki and of Stainbrook were emphasized. Could they not suggest something such as the following? It is reasonable to suppose that, immediately following a convulsion, the synaptic connections within the brain are weakened, distended. A virtually physical increase in ego permeability would thus occur. At the least, one must admit that individuals having just experienced a severe convulsion would be dazed and would have less control of affective drives; would of necessity respond to the environment and perceive it in a rather gross, undifferentiated fashion.

Brain-injured individuals. Where actual destruction of brain tissue is known to have occurred, most investigators (53, 64, 70, 78, 91) find that on the Rorschach test "organics" present a picture of extratensiveness, with $CF+C$ responses predominating over FC. There is an additional indication that may be present where the other is not—color naming. The functions that color naming may play in the dynamics of the child were previously discussed: color naming may be a rather concrete response evoked by the practice of parents teaching their children the names of various colors by pointing to an object and simply giving the name of the surface color; or it might be due to an attempt on the part of the child to designate an emotional situation without being able to act on his interpretation. It is doubtful which, if either, of these interpretations is applicable to the brain-injured individual. If it is the former, this is an indication of the adoption of an ineffective concrete attitude. If it is the latter, a bewilderment, an inability to integrate and resolve affective situations, may be indicated.

A further contribution. Bychowski (22) has published a paper resulting from his observations and study of individuals undergoing treatment and training in a rehabilitation center

for the brain injured. He found that, while certain physiological changes resulting from the injury might not be alleviated, a considerable improvement in the psychological behavior of the individual frequently resulted from the retraining given by the clinic. It is as if the adaptive and emotional skills of the individual had been destroyed or eliminated at least temporarily by the injury. It was necessary for the individual to re-learn, under very careful guidance and supervision, adaptive and emotional skills. The writer feels that Bychowski's report tends to confirm the postulates made in other sections of this paper about the relation between the affective picture presented by the individual and his capacity and opportunity to learn.

A Free Behavior Situation

Young and Higginbotham (96) attempted to correlate certain Rorschach factors with actual behavior in a free situation (summer camp). While a number of records corresponded approximately to the behavior in which the boys indulged, there were some striking exceptions —such as the child who was seemingly the most excitable and impulsive in the entire camp. This child had given no color responses on the Rorschach test, and the authors were inclined to question whether color was actually an indication of impulsivity; if it were, one might certainly expect this child to use a great deal of color, with most of it being of the *CF* or *C* variety. The authors' point was well taken. The present writer has expressed the opinion that impulsivity is not necessarily a concomitant of unbalanced color. Schachtel (77) suggested that affects (and color) are conscious manifestations of instinctual and other drives. An obvious conclusion

in the case of the particular individual in question is that he has repressed his affective drives and is not even attempting to handle them on a conscious basis or level, or that he has never learned to recognize affective behavior in himself or others. From such an individual one would expect quite flighty and "impulsive" actions.

The Strongest Challenge

Siipola, in a highly original experiment concerned with the effect of color upon the responses of individuals taking the Rorschach test, concluded: "Apparently, the mere presence of color in a blot does not endow it with magic affect-arousing properties" (81, p. 381). She did find, however, an increase in the number of emotional attitudes when the individual was confronted by the colored blots and, as a result of the color, "a weak selective influence among form dominated concepts, and a strong disruptive influence involving symptoms suggestive of conceptual conflict and behavioral disorganization" (81, p. 381).

Many individuals when confronted with an achromatic version of the blot gave exactly the same response given by those individuals confronted with the chromatic version. One example of such behavior was the response of "butterfly" to the middle red area of Card III. The present writer has found time and time again that some individuals when confronted with this blot in the chromatic version will give the response of butterfly, and then when asked what contributed to the forming of the concept answer that it was only the form, just the shape. The color was actually ignored. There are many other individuals who do respond to the color of the blot and use the color in developing their con-

cepts. It is the *response* to color which is important, and not just the fact that the color is there to be responded to by any except the color-blind individual.

Siipola (81) also found that those blots which especially aroused emotional responses or emotionally charged reactions were those in which the color was particularly incongruous to the usual forms that were conceived in the achromatic version. She felt that it was this incongruity of color and form which aroused the emotional reaction rather than simply the presence of color, for where there were no incongruities the emotional reactions did not occur. This finding ties in closely with her certainly well-grounded statement that as yet no one has succeeded in bridging the gap between color and affect by other than empirical data or theory based upon empirical data.

This last finding, that of emotional reactions being the product of color and form incongruity, may force a reconceptualization of the problem. Let it be granted, for the moment, that color in and of itself has no effect. Its mere presence is of no importance. However, some individuals will respond to the colored inkblot by incorporating the color into a concept. There are other individuals who will look at this same colored inkblot without incorporating the color into a concept. The point that affect-color theorists and empiricists wish to make is that the individual who does respond to color is in some way different from the individual who does not do so. Further, these theorists and empiricists feel that the manner in which the color is used is a very keen tool for analyzing the individual's affective life, and for gaining some insight into his ego functioning. At the present time, more

information exists upon the latter phase (as this review shows) than the former. ·A suggestion is that those individuals who do respond to color show more permeability, more susceptibility to influence by the environmental configuration than those individuals responding little or not at all to color.

Can it be granted that the mere presence of color has no effect upon the organism? Goldstein's well-known experiment with brain-injured individuals (37) seems to indicate clearly that exposing color to such individuals may markedly influence their physiological functioning. A study by Baccino (7) rather definitely indicates that chromatic illumination has a rather profound influence on the physiological functioning and growth of certain animal organisms. Conversely, the suggestion is strong that this strong response to color is actually due to definite brain damage resulting in decreasing effectiveness of inhibitory centers (91).

An experiment by Kravkov is highly provocative. This investigator injected into humans certain drugs which were known to have an effect upon the autonomic nervous system. He then compared the sensitivity of the eye to various colors.

"Changes in color sensation indicate a definite regularity depending upon the portion of the vegetative nervous system which is chiefly stimulated. Thus (use of) sympathetic toxins . . . bring about an increase in color sensitivity with respect to the green-blue rays of the spectrum and in contrast lower the color sensitivity with respect to the orange-red rays. The utilization of parasympathetic toxins brings about an increase in the sensitivity to orange-red rays and lowers the sensitivity to green-blue rays" (56, p. 94; translated from the Russian).

While these particular findings are not of tremendous importance here, the writer feels that the fact that alterations in the autonomic nervous system can cause a change in the behavior of individuals towards color is of great importance. Recently, there has been accumulating a considerable body of evidence to suggest that the condition of the autonomic nervous system may have profound effect upon behavior (31, 89).

A RETURN TO THE THEORY

There has been sketched in this paper a theory of the relation of the response to color and personality dynamics. Two papers, one by Schachtel (77), the other by Rickers-Ovsiankina (74), contributed heavily to its formulation. The theory with some of its ramifications is briefly stated below.

The response to color is indicative of a certain functioning of the ego: the relation of the ego to its external environment, its degree of communication with, and readiness to respond to it. One may determine much concerning the affective life of an individual through an analysis of his response to color: his control of affective charges, his capacity to interpret and integrate the affective behavior of others. But Rickers-Ovsiankina (74) speaks of the permeability of the ego. Does this permeability refer solely to affect or could it also be extended to cover intellective functions as well? Is there actually a rigid demarcation between an inividual's affective and intellective life areas? Can an individual be shut off affectively from his environment and still participate intellectively with it?

One of the personality patterns presented by Beck (12) may shed illumination. Beck draws a picture of a university president, a skilled scientist who has made valuable contributions in many areas. This individual gave nine *FC* and five *CF* responses. Interpreted according to the theory presented here, this individual is exceptionally responsive to the environment in which he lives, but he also has tremendous capacity and power to integrate and control this responsiveness.

Whence comes the material with which we create but the environment? If the degree of communication with the environment is limited, then the material with which to create is limited. An individual who has limited communication with his environment may be able to do much with what he has. But does not the individual with considerable environmental communication, granted the capacity to control and integrate his responsiveness, have a tremendous advantage over his fellowman who has not this responsiveness?

A second ramification concerns the clinician or other investigator seeking to learn of the affective life of an individual and his environmental responsiveness. A point made by Schachtel (77) must be continually kept in mind. Affect, and color, is the conscious manifestation of instinctual and other drives. If the individual has no or little experience of affect or of color, the possible inferences are three: (*a*) he may have an inherently limited capacity for affective experience, (*b*) he may be repressing affective experience, or (*c*) he may never have learned to express his affective drives. If the second condition exists, then the individual is not dealing with his affective life on a conscious level, and one should not expect to find representations of affective life, such as the use of color, in tests such as the Rorschach and in similar situations. This condition may have been operative in the case

of the "impulsive" individual described by Young and Higginbotham (96). If the third condition exists, one should anticipate evidence of limited and perhaps stereotyped affective experience. This third condition the writer feels to be particularly exemplified by the juvenile delinquent (69), although the second condition could also be operative. The selection by the clinician of the particular condition which is in effect must rest upon an analysis of other factors and a careful case history.

A caution must be made. For the affect-color theory to be functionally correct, there need not be any emotional or affective reaction to color. It is the *response* to color which is important. There need be no "magic affect-arousing properties" of color. There may be affect-arousing properties and emotional or affective reactions to color may occur, but these factors are not ingredients of this theory.

SUMMARY

1. A theory concerning the nature of the relation of the response to color and personality dynamics was presented. The theory strongly suggests that much can be learned from the response to color by the individual concerning the nature of the relation of the ego to the external environment as well as the relation of the ego to the affective drives of the individual.

2. A number of studies, drawn principally from the large body of Rorschach data, but also including several based upon the Mosaic Test and easel painting, were reviewed. The writer feels, first, that the theory is substantiated by the articles reviewed, and second, that a clarification of the dynamics of certain normal as well as disease processes results when the theory presented here is used in interpretation rather than prevalent practices.

BIBLIOGRAPHY

1. ABEL, T. M. The Rorschach test and school success among mental defectives. *Rorsch. Res. Exch.*, 1945, 9, 105–110.
2. AITA, J. A., *et al.* Rorschach's test as a diagnostic aid to brain injury. *Amer. J. Psychiat.*, 1947, 103, 770–780.
3. ALSCHULER, R., & HATTWICK, L. *Painting and personality; a study of young children.* Chicago: Univer. of Chicago Press, 1947.
4. ANASTASI, ANNE. An experimental study of the drawing behavior of adult psychotics in comparison with that of a normal control group. *J. exp. Psychol.*, 1944, 34, 169–194.
5. ARLOW, J., & KADIS, A. Fingerpainting in the psychotherapy of children. *Amer. J. Orthopsychiat.*, 1946, 16, 134–146.
6. ARLUCK, E. W. A study of some personality differences between epileptics and normals. *Rorsch. Res. Exch.*, 1940, 4, 154–156.
7. BACCINO, M. Étude de l'influence de la lumière colorée sur la croisance des jeunes homéothermes (lapereaux, cobayes, et rats blanc). *Séance du 25 Juin, 1938. Comptes rendus Hebdomadanes de la societé de biologie et de ses filiales,* 1938, 11, 767.
8. BAKER, L., & HARRIS, J. The validation of Rorschach test results against laboratory behavior. *J. clin. Psychol.*, 1949, 5, 161–164.
9. BECK, S. Personality structure in schizophrenics. *Nerv. ment. Dis. Monogr.*, 1938, No. 63.
10. BECK, S. The Rorschach test in psychotherapy. *J. consult. Psychol.*, 1943, 7, 103–111.
11. BECK, S. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune and Stratton, 1944.
12. BECK, S. *Rorschach's test.* Vol. II. *A variety of personality pictures.* New York: Grune and Stratton, 1947.
13. BECK, S., *et. al.* The normal personality as projected by the Rorschach test. *J. Psychol.*, 1950, 30, 241–298.
14. BILLIG, O., & SULLIVAN, D. Prognostic

data in chronic alcoholism. *Rorsch. Res. Exch.*, 1942, **6**, 117–127.

15. BLUM, L., & DRAGOWITZ, A. Fingerpainting: the developmental aspects. *Child Develpm.*, 1947, **18**, 88–105.

16. BOCHNER, R., & HALPERN, F. *The clinical application of the Rorschach test.* New York: Grune and Stratton, 1942.

17. BOWLUS, D., Shotwell, A. A Rorschach study psychopathic delinquency. *Amer. J. ment. Def.*, 1947, **52**, 23–30.

18. BOYNTON, P., & WALSWORTH, B. Emotionality test scores of delinquent and non-delinquent girls. *J. abnorm. soc. Psychol.*, 1943, **38**, 87–92.

19. BRENMAN, MARGARET, & REICHARD, S. Use of the Rorschach test in the prediction of hypnotizability. *Bull. Menninger Clin.*, 1943, **7**, 183–187.

20. BRUSSEL, J., *et al.* The Rorschach method and post concussion syndrome. *Psychiat. Quart.*, 1942, **16**, 707–743.

21. BUHLER, CHARLOTTE. The concept of integration and the Rorschach test as a measurement of personality integration. *J. proj. Tech.*, 1950, **14**, 315–319.

22. BYCHOWSKI, G. The ego of the brain-wounded. *Psychoanal. Rev.*, 1949, **36**, 333–343.

23. CLARK, J. Some MMPI correlates of color responses in the Rorschach. *J. consult. Psychol.*, 1948, **12**, 384–386.

24. COLM, H. The value of projective methods in the psychological examination of children. The Mosaic test in conjunction with the Rorschach and Binet tests. *Rorsch. Res. Exch.*, 1948, **12**, 216–237.

25. DAVIDSON, H., & KLOPFER, B. Rorschach statistics: I. Mentally retarded, normal and superior adults. *Rorsch. Res. Exch.*, 1938, **2**, 164–169.

26. DIAMOND, B., & SCHMALE, H. The Mosaic test. I. An evaluation of its clinical application. *Amer. J. Orthopsychiat.*, 1944, **14**, 237–251.

27. ELLIOT, M., & MERRILL, F. *Social disorganization.* New York: Harper, 1941.

28. ENDACOTT, J. L. The results of 100 male juvenile delinquents on the Rorschach ink-blot test. *J. crim. Psychopathol.*, 1941, **3**, 41–50.

29. EPSTEIN, H., & SCHWARTZ, A. Psychodiagnostic testing in group work (Rorschach and painting analysis technique). *Rorsch. Res. Exch.*, 1947, **11**, 23–41.

30. FORD, M. The application of the Rorschach test to young children. *Univer. Minn. Child Welf. Monogr.*, 1946, No. 23.

31. GELLHORN, E. *Autonomic regulations: their significance for physiology, psychology and neuropsychiatry.* New York: Interscience Publishers, 1942.

32. GESELL, A. L., & ILG, F. *Infant and child in the culture of today.* New York: Harper, 1943.

33. GOLDFARB, W. Personality trends in a group of enuretic children below the age of 10. *Rorsch. Res. Exch.*, 1942, **6**, 28–36.

34. GOLDFARB, W. Effects of early institutional care on adolescent personality: Rorschach data. *Amer. J. Orthopsychiat.*, 1944, **14**, 441–447.

35. GOLDFARB, W. Rorschach test differences between family-reared and institution-reared, and schizophrenic children. *Amer. J. Orthopsychiat.*, 1949, **19**, 624–633.

36. GOLDFARB, W., & KLOPFER, B. Rorschach characteristics of "institution" children. *Rorsch. Res. Exch.*, 1944, **8**, 92–100.

37. GOLDSTEIN, K. *The organism.* New York: American Book Co., 1939.

38. GUIRDHAM, A. The Rorschach test in epileptics. *J. ment. Sci.*, 1935, **81**, 870–893.

39. HAMLIN, R., *et al.* Objective Rorschach signs for groups of normal, maladjusted and neuropsychiatric subjects. *J. consult. Psychol.*, 1950, **14**, 276–282.

40. HARTMAN, H., *et al.* Comments on the formation of psychic structure. *Psychoanal. Stud. Child*, 1946, **2**, 11–38.

41. HERTZ, M. R. Personality patterns in adolescence as portrayed by the Rorschach ink-blot method: III. The "Erlebnistypus" (a normative study). *J. gen. Psychol.*, 1943, **28**, 225–276.

42. HERTZ, M. R., & BAKER, E. Personality patterns in adolescence as portrayed by the Rorschach ink-blot method: II. The color factors. *J. gen. Psychol.*, 1943, **28**, 3–61.

43. HERTZMAN, M., & MARGUILES, H. Developmental changes as reflected in Rorschach test responses. *J. genet. Psychol.*, 1943, **62**, 189–215.

44. HEUSER, K. D. The psychopathic personality: Rorschach patterns of 28 cases. *Amer. J. Psychiat.*, 1946, **103**, 105–112.

45. JERSILD, A. Emotional development. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946.

46. KATES, S. Objective Rorschach response patterns differentiating anxiety reactions from obsessive-compulsive reac-

tions. *J. consult. Psychol.*, 1950, **14**, 226–229.

47. KELLEY, D. M. Intravenous sodium amytal medication as an aid to the Rorschach method. *Psychiat. Quart.*, 1941, **15**, 68.

48. KELLEY, D. M., & BARRERA, S. The Rorschach method in the study of mental deficiency. *Amer. J. ment. Def.*, 1941, **45**, 401–407.

49. KELLEY, D. M., & KLOPFER, B. Application of the Rorschach method to research in schizophrenia. *Rorsch. Res. Exch.*, 1938, **3**, 55–65.

50. KELLEY, D. M., & MARGUILES, H. Rorschach case studies in the convulsive states. *Rorsch. Res. Exch.*, 1940, **4**, 157–189.

51. KENDIG, I. Rorschach indications for the diagnosis of schizophrenia. *J. proj. Tech.*, 1949, **13**, 142–149.

52. KISKER, G. A projective approach to personality patterns during insulin-shock and metrazol therapy. *J. abnorm. soc. Psychol.*, 1942, **37**, 120.

53. KLEBANOFF, I. Psychological changes in organic brain lesions and ablations. *Psychol. Bull.*, 1945, **42**, 585–623.

54. KLOPFER, B., & KELLEY, D. *The Rorschach technique: a manual for a projective method of personality diagnosis.* Yonkers-On-Hudson, New York: World Book Co., 1946.

55. KLOPFER, B., & MARGUILES, H. Rorschach reactions in early childhood. *Rorsch. Res. Exch.*, 1941, **5**, 1–23.

56. KRAVKOV, S. O sviaziakh tsvetnogo zreniia s vegetativnoi nervnoi sistenoi. *Probl. physiol. Opt., Acad. Sci., U.S.S.R.*, 1941, **1**, 87–95.

57. LEVINE, K., *et al.* Hypnotically induced mood changes in the verbal and graphic Rorschach. A case study. *Rorsch. Res. Exch.*, 1944, **8**, 104–124.

58. LINDNER, R. The Rorschach test and the diagnosis of psychopathic personality. *J. crim. Psychopathol.*, 1943, **5**, 69–93.

59. LINN, L. The Rorschach test results in the evaluation of military personnel. *Rorsch. Res. Exch.*, 1946, **10**, 20–27.

60. LOWENFIELD, M. The Mosaic test. *Amer. J. Orthopsychiat.*, 1949, **19**, 537–550.

61. McCULLOCH, T., & GIRDNER, J. Use of the Lowenfield Mosaic test with mental defectives. *Amer. J. ment. Def.*, 1949, **53**, 486–496.

62. McFATE, M., & ORR, F. Through adolescence with the Rorschach. *Rorsch. Res. Exch.*, 1949, **13**, 302–319.

63. McLEOD, H. A Rorschach study with pre-school children. *J. proj. Tech.*, 1950, **14**, 453–464.

64. NADEL, A. A qualitative analysis of behavior following cerebral lesions diagnosed as primarily affecting the frontal lobes. *Arch. Psychol.*, 1938, **32**, No. 224.

65. NAPOLI, P. Fingerpainting and personality diagnosis. *Genet. Psychol. Monogr.*, 1946, **34**, 129–230.

66. NAPOLI, P. Interpretative aspects of fingerpainting. *J. Psychol.*, 1947, **23**, 93–132.

67. OESER, O. Some experiments on the abstraction of form and color: II. Rorschach tests. *Brit. J. Psychol.*, 1932, **22**, 287–323.

68. PEMBERTON, W. General semantics and the Rorschach test. *Papers from the Second American Congress on general semantics.* Chicago, 1943.

69. PHILLIPS, E., & STROMBERG, E. A comparative study of fingerpainting performance in detention homes and high school pupils. *J. Psychol.*, 1948, **26**, 507–515.

70. PIOTROWSKI, Z. The Rorschach ink-blot method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525–537.

71. PRADOS, M., & FRIED, E. Personality structures of the older age groups. *J. clin. Psychol.*, 1947, **3**, 114–120.

72. PRECKER, J. A. Painting and drawing in personality assessment. *J. proj. Tech.*, 1950, **14**, 262–286.

73. RICKERS-OVSIANKINA, MARIA. The Rorschach test as applied to normal and schizophrenic subjects. *Brit. J. med. Psychol.*, 1938, **17**, 227–257.

74. RICKERS-OVSIANKINA, MARIA. Some theoretical considerations regarding the Rorschach method. *Rorsch. Res. Exch.*, 1943, **7**, 41–53.

75. RORSCHACH, H. *Psychodiagnostics.* New York: Grune and Stratton, 1949.

76. SARASON, S., & POTTER, E. Color in the Rorschach and Kohs block designs. *J. consult. Psychol.*, 1947, **11**, 202–206.

77. SCHACHTEL, E. On color and affect; contributions to an understanding of Rorschach's test. II. *Psychiatry*, 1943, **6**, 393–409.

78. SCHAFER, R. *The clinical application of psychological tests.* New York: International Universities Press, 1949.

79. SCHMIDL, F. The Rorschach test in delinquency research. *Amer. J. Orthopsychiat.*, 1947, **17**, 151–160.

80. SIEGEL, M. The diagnostic and prognostic validity of the Rorschach test in a

child guidance clinic. *Amer. J. Orthopsychiat.*, 1948, **18**, 119–133.

81. SIPOLA, E. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.

82. STAINBROOK, E. The Rorschach description of immediate post-convulsive mental function. *Charact. & Pers.*, 1943–1944, **12**, 302–322.

83. STEIN, M. Personality factors involved in the temporal development of Rorschach responses. *Rorsch. Res. Exch.*, 1949, **13**, 355–414.

84. STEINZOR, B. Rorschach responses of achieving and non-achieving college students of high ability. *Amer. J. Orthopsychiat.*, 1944, **14**, 494–504.

85. STERN, K., & MALLOY, H. Rorschach studies on patients with paranoid features. *J. clin. Psychol.*, 1945, **1**, 272–280.

86. SWIFT, J. Rorschach responses of 82 preschool children. *Rorsch. Res. Exch.*, 1945, **9**, 74–84.

87. WAEHNER, T. Interpretation of spontaneous drawings and paintings. *Genet. Psychol. Monogr.*, 1946, **33**, 3–70.

88. WARREN, H. *Dictionary of psychology.* Boston: Houghton Mifflin, 1934.

89. WENGER, M. Preliminary study of the significance of measures of autonomic balance. *Psychosom. Med.*, 1947, **9**, 301–309.

90. WERNER, H. Rorschach method applied to two clinical groups of mental defectives. *Amer. J. ment. Def.*, 1945, **49**, 304–306.

91. WERNER, H. Perceptual behavior of brain injured children: an experimental study by means of the Rorschach test. *Genet. Psychol. Monogr.*, 1945, **31**, 51–110.

92. WERTHAM, F., & GOLDEN, L. A differential diagnostic method of interpreting the Mosaic and colored block designs. *Amer. J. Orthopsychiat.*, 1941, **98**, 124–131.

93. WILKINS, W., & ADAMS, A. The use of the Rorschach test under hypnosis and under sodium amytal in military psychology. *J. gen. Psychol.*, 1947, **36**, 131–138.

94. WILLIAMS, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.

95. WITTENBORN, J. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261–267.

96. YOUNG, R., & HIGGINBOTHAM, S. Behavior checks on the Rorschach method. *Amer. J. Orthopsychiat.*, 1942, **12**, 87–94.

BAUER, RAYMOND A. *The new man in Soviet psychology.* Foreword by Jerome S. Bruner. Cambridge, Mass.: Harvard Univer. Press, 1952. Pp. xxiii+229. $4.00.

In view of the very real curiosity—to use a mild term—which most of us feel regarding events on the other side of the semipermeable (?) membrane which separates us from the Russians, this little volume should be warmly welcomed by American psychologists. Since so few of us have direct access, for various reasons, to the Soviet psychological literature, and since contacts with our Russian colleagues have been reduced to the vanishing point, we are all indebted to Bauer for bringing us relatively up to date on recent developments.

From the point of view of scientific psychology the picture which Bauer presents is generally discouraging. No journal, specifically psychological in nature, has appeared since 1934; articles of psychological interest are published mainly in a journal devoted to pedagogy. There has been a strong reaction, still evident, against psychology as an independent discipline. The study of attitudes has been condemned and virtually abandoned. No public opinion surveys may be conducted, and social psychology in general "has become virtually a proscribed area" (p. 169). A Party decree in 1936 resulted in the almost complete suppression of the use of psychological tests, which were "formally characterized as instruments for perpetuating the class structure of bourgeois societies" (p. 124). Scientific theory, in psychology as elsewhere, is validated not in terms of its relation to empirically verified facts, but by the contribution it can make

to the Party's program. A psychologist is quoted as stating that "every theoretical mistake, every error in the field of methodology is inescapably transferred into a political error" (p. 106). It goes without saying that "incorrect" views cannot be expressed or tolerated.

Bauer characterizes his book as "partially a history of the science of psychology in the Soviet Union, partially a study of the pattern of social change in that country, largely an analysis of changing conceptions of human nature under conditions of social change, to a certain extent an inquiry in the relation of ideology to action, somewhat a study of the relationship of psychology to society" (p. ix). On the whole this ambitious program is effectively realized. There is, of course, a close relation between the character and structure of a society and the current beliefs concerning the nature of man, his goals and motives, his development and socialization. Bauer has presented some striking correspondences between the character of the society as a whole, and the specific developments that have occurred in the field of psychology.

Nowhere does this come out more clearly than in his discussion of the changes which have occurred in the concept of the nature of man, as a reflection of the changes which took place in the political scene. In the 1920's, for example, Soviet psychologists proceeded on the assumption that man's nature was essentially passive, his characteristics determined by the (mainly economic) environment. Fundamentally good and noble, man had been misled and perverted by the evil (that is, bourgeois,

capitalist) system under which he lived. Even after the Bolshevik revolution, the environment could still be held responsible, because it contained many of the elements of the older social and economic structure. By 1936, however, socialism had allegedly been fully realized in the Soviet Union. From that time on, it became impossible to blame the (socialist) environment; the responsibility was now placed upon the individual himself. "The dominant conception of man became that of an increasingly purposeful being, who was more and more the master of his own fate, and less and less the creature of his environment" (p. 7).

That meant a movement away from behaviorism, reflexology, and "reactology" to a more "purposive" variety of psychology. Consciousness was restored to a dominant role in human affairs, and the unconscious fell correspondingly into disfavor; not that its existence was denied, but rather that it became subordinate in importance to conscious, purposive action. The source of error was now found at least in part in man himself; he needed the right training, and the right self-training, to set him upon the proper road. From the point of view of Western psychology, training is of course a part of what we would call the environment. Bauer suggests that this distinction is drawn by Soviet psychologists "mainly to deprecate the importance of such aspects of the environment as the actual material conditions under which the child lives" (p. 148). The question arises as to whether the words for "environment" have a somewhat different connotation in Russian and English respectively, in which case the Soviet attack on environmental explanations of behavior would really represent an attack on a very restricted variety of environ-

mentalism. In any case, the coincidence between the political pronouncements of 1936 and the attack of the prevailing science of psychology is a striking one.

Bauer draws an interesting contrast between the Soviet view of man and that prevalent in Nazi Germany. The Nazi view held that man was moved primarily by the unconscious, the nonrational; the Bolshevik stresses consciousness and rationality. The Nazi stressed man's weakness, his helplessness, his need of a leader; the Bolsheviks insist on man's responsibility for his behavior, on his ability to make his own destiny—though the only "right" destiny is that which follows the party line. "For the Nazi, man was a marionette who moved when one pulled the strings. For the Bolshevik, he is a robot who can be trained to act independently within specified limits" (p. 178). One may argue about some of the details of these characterizations, but it seems clear that the two dictatorships do differ markedly in the meanings they attach to human nature. Perhaps we have here a clue to a differential diagnosis of varieties of dictatorship in psychological terms.

Bauer's informative study would have been still more valuable, at least in the opinion of this reviewer, if he had included somewhat more discussion of some of the specific investigations carried out by Russian psychologists. Granting that his interest was in the development of theory, his thesis could have been more clearly illuminated by a fuller demonstration of the manner and extent to which theory dominated the collection of "facts." One might argue also about the amount of emphasis which Bauer places on the discontinuities to be found in Soviet psychology. It is interesting that Dr. Joseph Wortis in his *Soviet Psychia-*

try (Baltimore, 1950), as Bauer himself indicates, was much more forcibly struck by the continuities. This difference in interpretation by two scholars examining closely related material calls for fuller exploration than that contained in a brief footnote. These are relatively minor issues, however, compared to the over-all value of Bauer's study. The Russian Research Center at Harvard has made a significant addition to its excellent series.

The last sentence of the book is worth repeating. "Political interference in science does not destroy completely the usefulness of science to the system, but the continued suppression of freedom of scientific inquiry must ultimately lead to the point where the society cannot solve its own problems effectively" (p. 196). This applies in the United States just as it does elsewhere. It cannot be said too often: there can be no real development of science where there is no freedom—freedom to explore, to doubt, to criticize, to deviate, even to be wrong. There are people in this country, too, who have taken it upon themselves to tell scientists, psychologists and others, what they may teach and what they may discover. Soviet psychology should serve us as an object lesson. If we allow that kind of interference here, we might just as well shut up shop.

<div align="right">Otto Klineberg.</div>
Columbia University.

Jaques, Elliott. *The changing culture of a factory.* New York: Dryden Press, 1952. Pp. xxi+341. $4.25.

This is the published report of ". . . a case study of developments in the social life of one industrial community between April, 1948 and November 1950." The "case" is a small, publicly held British company engaged principally in the manufacture, sale, and servicing of metal bearings. The study is concerned with the description, diagnosis, and treatment of the corporate syntality. The results reported are the product of the collaborative efforts of the personnel of the company and of a thirteen-member research team headed by the author of the book, Dr. Jaques. The research was sponsored by the Tavistock Institute of Human Relations, London, and has been accepted as a Ph.D. thesis in the Department of Social Relations at Harvard.

The first part provides a retrospective glimpse of the corporate organization as it evolved during the first fifty years of life. This is followed by the case study proper, a detailed description of events as they occurred during the period of observation. The methods used by the research team to gain acceptance by company personnel at all levels and to function effectively in the multiple role of consultant, analyst, and therapist are a highlight of this second section of the report. The third and concluding part of the book contains an analysis and interpretation of the findings. "The method of analysis . . . will be to study how the pattern of social activity at Glacier (firm name) . . . has come about through the interaction of the firm's organizational structure, its customary way of doing things, and the behavior of its members . . . we shall study the interaction of social structure, culture, and personality." The results presented in this part point up the need for defining and clarifying individual and group roles as an antecedent step both to understanding social behavior and to evaluating it. Inferentially, the adequacy of a group's adjustment, in large measure, is considered to be a function of the members' understanding of the au-

thority-responsibility relationship and possession of authority by the individual members commensurate with their felt responsibilities.

The study exemplifies social science at its best, transcending the boundaries of any single professional research area. From the standpoint of both content and methodology it warrants the attention of all psychologists concerned with interpersonal and intergroup relations as they affect the individual's adjustment. Generalization from the specific findings is, of course, limited by the very nature of the case study method and, in this instance, by the concurrent diagnosis and counseling required of the investigators during the course of the study.

WILLIAM J. E. CRISSY.
Queens College.

JUDD, DEANE B. *Color in business, science and industry.* New York: Wiley, 1952. Pp. ix+401. $6.50.

The author of *Color in Business, Science and Industry* seems to address himself primarily to business men and industrialists to call attention to the scientific aids now available for the solution of a variety of practical color problems. In the preface he says, "It has been my privilege . . . in my twenty years at the National Bureau of Standards, to come into contact with hundreds of colorimetric sore spots in our industrial life. I have seen victories that paid off in dollars and cents won by applying the sciences of mathematics, physics, and psychology to these problems." The author refers specifically to a great number of practical color problems encountered in everyday life and endeavors to indicate how "visual psychophysics mixed with a liberal sprinkling of common sense" can provide a solution for these problems.

The work is divided into three principal sections: Part I, Basic Facts; Part II, Tools and Technics; and Part III, Physics and Psychophysics of Colorant Layers. Cursory inspection reveals that Part II is the most important. More than half the book is devoted to this part, which is nearly three times as long as Part I and four times as long as Part III.

Part I lays the groundwork for the later exposition. It includes a twenty-page treatment on the structure and functions of the eye, a summary of the basic physical, psychological, and psychophysical terms currently employed in the field of color perception, a discussion of methods of color matching (a) by addition of lights, (b) by rapid succession of lights, and (c) by mixture of colorants. The first part closes with a discussion of different types of color deficiency and a brief description of the better known tests of color blindness.

The various tools and technics used by the color specialist are described in Part II. The reader is informed in some detail, with the aid of many diagrams and with the necessary quantitative tables, concerning spectrophotometers, the standard observer, chromaticity diagrams, tristimulus values and tristimulus colorimeters, subtractive colorimeters, photometers, photoelectric tristimulus colorimeters, color standards, color scales, and color names. How each of these aids is to be applied in connection with diverse manufacturing problems is set forth in an easy running style with great clarity. Here the author has rendered important service to workers in the color practicum, not only by indicating what each of these aids is designed to do and how it is to be applied, but also by pointing out some of their limitations.

Part III is devoted to special problems, such as gloss, opacity or hiding power, clear and turbid media, which

have been subjected to intensive quantitative analyses. The solutions of these problems, sometimes involving exponential and hyperbolic functions, will be of interest chiefly to those of high mathematical competence.

About fifteen pages are devoted to each of the three final sections of the work: (*a*) an appendix, containing quantitative tables too extended to be included in the main body of the text, (*b*) a list of references, and (*c*) the index. Although the list of references contains more than three hundred and fifty items, a number of very important contributors to color are omitted. There is no reference to E. Hering, G. E. Müller, L. T. Troland, C. E. Ferree, and S. Hecht, to mention only a few of those no longer living.

Of the several sciences sharing in the scientific study of color, physics and psychology fare somewhat better than physiology. However, the opening treatment in Part I is designed to provide some balance among the several sciences with overlapping interests in color. Despite the attempted integration among these sciences, it seems fair to say that Judd's chief contribution consists in the description of the methods by which the physical correlates of the visual stimulus are to be specified. What seems to be left for other men of science, perhaps in electrophysiology, is the task of surveying more fully the possibilities of specifying and, in so far as possible, of rendering to quantitative terms the physiological correlates of the visual stimulus.

MICHAEL J. ZIGLER.
Wellesley College.

KARN, HARRY W., AND GILMER, B. VON HALLER. *Readings in industrial and business psychology.* New York: McGraw-Hill, 1952. Pp. ix+476. $4.50.

The trend toward books of readings in various areas of psychology is continued with the publication of this volume. Some such books have tended to present materials from all eras of psychology, some have had the original articles edited rather sharply, and some have included extensive comments by the editors directed toward emphasis and integration. The present volume does none of these. The editors state that they have not attempted integration nor have they tried to cover articles of historical interest. They have tried to include "representative" articles which are easily understood by those lacking extensive technical training.

The aim of recency is realized, 41 of the 53 articles having been published since 1944. The articles are drawn from 19 different sources with approximately 25 per cent of the articles having been published originally in either the *Journal of Applied Psychology* or *Personnel Psychology.* The articles are rather evenly spread among 11 "fields" of business and industrial psychology. There is some question regarding "representativeness" in the selected articles. For example, in terms of content, the editors might have selected a more representative article on sampling in market research than the one by Stanton written in 1941 which makes no reference to probability sampling. Furthermore, the number of articles devoted to each of the fields is not proportionate to the amount of research or information in those fields. Although this reviewer is not opposed to disproportionate sampling of articles, the term representative is hardly used in its usual sense and should be read as illustrative. It is good to see sections devoted to the place of the psychologist in industry and to his ethical problems. The

comments made by the editors about each article are extremely brief and of no great value. These comments could have been expanded and might have included a few words about each author.

As far as this reviewer knows, this is the first attempt to provide a book of general readings in business and industrial psychology since Moore and Hartmann's *Readings* of 20 years ago. The latter largely contained excerpts, whereas Karn and Gilmer present complete articles. Direct comparison of the books is probably unfair since the selections were not made on the same bases in the two books. However, a casual comparison shows that the present volume devotes more space to training, counseling, job evaluation, market research, fatigue and efficiency, and leadership, and much less to tests and selection than did the earlier book. About the same amount of space is devoted to motivation and morale, and industrial relations.

As the editors point out, this book should be a supplement to a systematic text or else the teacher must organize and integrate the subject matter. The main advantage to this book is that a number of recent, illustrative articles are combined under one cover.

LESTER GUEST.
The Pennsylvania State College.

YOUNG, KIMBALL. *Personality and problems of adjustment.* (2nd Ed.) New York: Appleton-Century-Crofts, 1952. Pp. x+716. $5.00.

A considerable amount has been added to the first edition of this very readable book, which appeared first in 1940 (see *Psychological Bulletin*, 1941, 188ff.). The author has included some fairly recent material from the perception-personality area, but the presentation is substantially eclectic. Every sort of approach has been covered, including George Mead's subjective analysis of the "I" and "Me," anthropological and psychoanalytical and field theories, and the personalistic contribution of Gordon Allport, leading into case studies. Some might object to Young's disinclination to take a point of view and stick to it. However, the average student will probably be benefited by the broad coverage.

The author is quite skeptical of our progress in the strictly experimental study of personality measurement. He states, "there are a large number of theories (of personality) but unfortunately they are not, for the most part, so stated as to furnish a bridge to empirical testing either in the laboratory or by other scientific devices."

Part I is mainly concerned with theories about personality and its development. There are traditional presentations of language and other forms of learning. Little has been added here, and the chapters on symbolic behavior and the self have apparently not been altered much. Both of these areas remain useful summaries of often neglected topics.

Part II, concerned with problems of adjustment, is the most concrete and will doubtless be read with major interest. As in the first edition, there are chapters devoted to infancy, childhood, adolescence, the college student, marriage, and neuroticism. There is in addition a completely new chapter on later maturity and old age, indicative of this rapidly expanding field. This reviewer feels that the chapter on psychological problems associated with occupation is given less up-to-date attention than it deserves, in comparison, for example, with the far greater consideration given to sexual and marital adjustment problems.

EDWARD S. JONES.
University of Buffalo.

THOMPSON, GEORGE G. *Child psychology: growth trends in psychological adjustment.* Boston: Houghton Mifflin, 1952. Pp. xxxiv+667. $5.50.

Oftentimes the child psychologist finds it difficult to convince himself or his colleagues that his area of interest has any specific contribution to make to psychology as a whole. In this book Thompson presents a convincing demonstration that when it is interpreted as the psychology of *development* (rather than as the psychology of children, to be contrasted with the psychology of apes or adolescents or old people) child psychology is a scientific discipline in its own right. This is a book developed thoughtfully, carefully, and with a high degree of scholarship. Obviously familiar in detail with the vast literature in developmental psychology, the author has been guided in his selection of material by criteria of scientific validity and pertinence of the material to the point he is trying to make. In other words, while the book is comprehensive it is not padded. Each of its fourteen chapters is organized as a unit and these units contribute to a functional whole.

Thompson's objective seems to have been to present an integrated picture of the developmental process underlying human behavior. Where necessary to round out the picture he has felt free to use findings with infrahuman "children," studies of individual cases, concepts from personality theory and the like, pointing out frequently that in many important areas of child psychology our knowledge is all too sparse. But pervading the whole is the insistence on the application of scientific principles in evaluating data. Considerations such as numbers of cases, the design of studies, and the reliability of observation are given repeated and consistent emphasis.

At various points in the book it becomes apparent that behavioral development does not always take place in a desirable direction, and at these points the author is particularly interested in examining any evidence which might explain such trends. This leads to careful consideration of methods of child rearing and guidance and education, in our own culture and in others.

In his general orientation, the author's position is not far from that of the modern behaviorist, but he has given considerable effort to the task of presenting fairly such materials as projective techniques and the theory underlying their use, if with an eclectic sort of damning with faint praise. Placing a premium upon factors of validity and reliability, it is natural that he would say of these techniques, "we should be cautious about our interpretations until more objective methods of scoring and interpretation have been worked out, and until more validation studies have been conducted with positive results" (p. 620). Exercising perhaps the same requirement of objective proof, the author gives almost no space at all to therapy as such.

This book will be well received by those who feel that child psychology should be taught as a science. It is written in a fashion sufficiently interesting to be appropriate for the non-psychology major who wants to learn something about children before generating his own, perhaps, and in a fashion sufficiently scholarly and comprehensive to satisfy the graduate student looking for a standard reference in this field. This book is to be highly recommended.

T. W. RICHARDS.
Louisiana State University.

BOOKS AND MONOGRAPHS RECEIVED

ASHBY, W. ROSS. *Design for a brain.* New York: Wiley, 1952. Pp. ix+259. $6.00.

ASCH, SOLOMON. *Social psychology.* New York: Prentice-Hall, 1952. Pp. xvi+646. $5.50.

BARLOW, FRED. *Mental prodigies.* New York: Philosophical Library, 1952. Pp. 256. $4.75.

BENDER, MORRIS B. *Disorders in perception.* Springfield, Ill.: Charles C Thomas, 1952. Pp. viii+109. $3.00.

BERGLER, EDMUND. *The superego.* New York: Grune & Stratton, 1952. Pp. x+367. $6.75.

BERNARD, HAROLD W. *Mental hygiene for classroom teachers.* New York: McGraw-Hill, 1952. Pp. xiii+472. $4.75.

BETT, W. R. *The infirmities of genius.* New York: Philosophical Library, 1952. Pp. 192. $4.75.

DAVIDSON, HENRY A. *Forensic psychiatry.* New York: Ronald Press, 1952. Pp. viii+398. $8.00.

DEESE, JAMES. *The psychology of learning.* New York: McGraw-Hill, 1952. Pp. ix+398. $5.50.

FERGUSON, LEONARD W. *Personality measurement.* New York: McGraw-Hill, 1952. Pp. xv+457. $6.00.

HIRSH, IRA J. *The measurement of hearing.* New York: McGraw-Hill, 1952. Pp. ix+364. $6.00.

JERSILD, ARTHUR T. *In search of self.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1952. Pp. xii+141. $2.75.

KERR, MADELINE. *Personality and conflict in Jamaica.* Liverpool: Univer. Press of Liverpool, 1952. Pp. xii+221. 15s.

LINDNER, ROBERT. *Prescription for rebellion.* New York: Rinehart, 1952. Pp. 305. $3.50.

MAGNE, OLOF. *Perception and learning.* Göteborg: Gumperts Förlag, 1952. Pp. 228. 25 kronor.

MASLOW, PAUL. *Powers of the mind.* Vol. II of the Life Science. Brooklyn: Author, 1952. Pp. 135–288. $3.50 (Mimeo).

McCULLOCH, WARREN S. *Finality and form.* Springfield, Ill.: Charles C Thomas, 1952. Pp. 63. $3.75.

MERTON, ROBERT K., FISKE, MARJORIE, & KENDALL, PATRICIA. *The focused interview.* New York: Bureau of Applied Social Research, Columbia Univer., 1952. Pp. xxv+202. $3.00.

MEYER, ALAN S. *Social and psychological factors in opiate addiction.* New York: Bureau of Applied Social Research, Columbia Univer., 1952. Pp. ix+170. $1.00.

RAUSCH, EDWIN. *Struktur und Metrik figural-optischer Wahrnehmung.* Frankfort am Main: Dr. Walter Kramer, 1952. Pp. xiv+404.

SMITH, GUDMUND. *Interpretations of behavior sequences.* København: Society of Psychology and Pedagogy, 1952. Pp. 120. 9:50 cr.

SWANSON, GUY E., NEWCOMB, THEODORE M., & HARTLEY, EUGENE L. *Readings in social psychology.* (Rev. Ed.) New York: Holt, 1952. Pp. xviii+680. $5.00.

VON BERTALANFFY, LUDWIG. *Problems of life.* New York: Wiley, 1952. Pp. xi+216. $4.00.

WALLS, GORDON L., & MATHEWS, RAVENNA W. *New means of studying color blindness and normal foveal color vision.* Berkeley: Univer. of California Press, 1952. Pp. iv+172. $2.50.

WEINLAND, JAMES D., & GROSS, MARGARET V. *Personnel interviewing.* New York: Ronald, 1952. Pp. vii+416. $6.00.

WITHERINGTON, H. CARL. *Educational psychology.* (Rev. Ed.) Boston: Ginn, 1952. Pp. v+487. $4.00.

Psychological Bulletin

MOTOKAWA'S STUDIES ON ELECTRIC EXCITATION OF THE HUMAN EYE[1]

J. W. GEBHARD

Applied Physics Laboratory
The Johns Hopkins University
Silver Spring, Maryland

Much progress in visual science has been made by measuring the electric response of the eye to a photic stimulus. In this connection, one need only point to the work of Granit (16), Hartline (18), and many others. Less deeply explored is the possibility of revealing some of the mechanisms of vision by using an electric input to the eye and measuring the response in terms of the sensation of light such a stimulus evokes. The discovery that an electric stimulus will easily produce a sensation of bluish-white light was made by Le Roy (30) in 1755, and this effect has interested physiologists and psychologists sporadically ever since. Visual perceptions caused by inadequate stimuli are called phosphenes and it was suspected long ago that the electrical phosphenes might be altered to some extent by photic stimulation. The first systematic work

on this problem appeared shortly after it became possible to control electric stimulation accurately. Electrical thresholds were measured at various levels of light adaptation and during the course of dark adaptation with conflicting results. Early investigators found no differences between the effects of light adaptation and dark adaptation (8, 28, 68), but later workers, such as Verrijp (81), Achelis and Merkulow (2), Bogoslovsky (3), and Schwarz (75), found the electrical threshold to be higher during dark adaptation. A new and perhaps very important turn has now been given to the problem of interaction between photic and electric stimulation by Prof. Kōiti Motokawa and his associates at the Physiological Laboratory of the Tohoku University, Sendai, Japan. These workers have been ingenious in their use of the threshold of an electrical phosphene as a means of analyzing the effect of photic stimulation on the activity of the visual mechanism. In a rather formidable output of some forty papers in the last four years, they have touched upon matters of brightness and color discrimination, color blindness, contrast, spatial summation, inhibition, flicker effects, visual illusions, and a new visual method of measuring fatigue. This is

73

10,000 lux.[3] Adaptation was followed for 40 min. with a Nagel adaptometer. The results are shown in Fig. 2 for both a 12° area (A) and a 1° area (B) with central fixation. The solid curve is for threshold sensitivity to light as normally measured. The dotted curve shows the results when

done: the electrical threshold for just noticeable flicker phosphenes at 14.3 cy./sec. was measured with, and without, threshold photic stimulation. Twenty-nine minutes of dark adaptation followed. The electrical thresholds continued to rise during the first 15 min. of dark adaptation
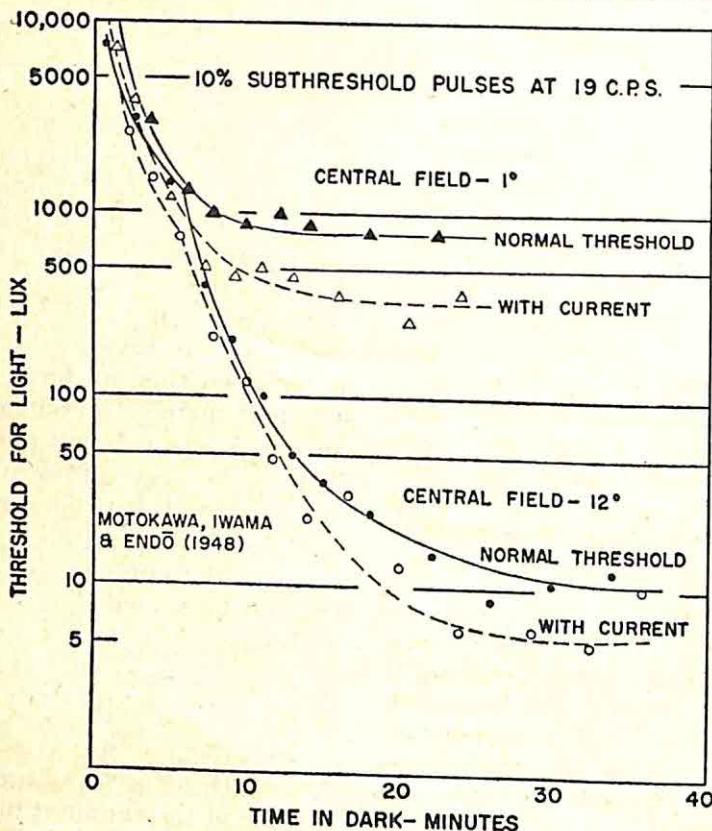

FIG. 2. EFFECT OF A SUBLIMINAL ELECTRIC STIMULUS ON THE PHOTIC THRESHOLD.
[AFTER MOTOKAWA, IWAMA, AND ENDŌ (62).]

the current pulses were used. The presence of an electric current is shown in the data to produce an eye that is more sensitive to light. This appears true over both the rod and the cone portions of the dark-adaptation curve.

Sensitizing Light

The reverse experiment was also

as measured both with and without light stimulation, and then leveled off. At all times, the electrical thresholds were lower in the presence of light.

This experiment clearly shows an interaction between electric and photic excitation processes and suggests that something gets sensitized by electric stimuli that makes it easier for light to act, and vice versa. However, it does not tell much about the site of electric stimulation. The

[3] The lux, or meter-candle, is equivalent to .0929 foot-candles. Ten thousand lux, therefore, are 929 foot-candles.

suggestion has frequently been made that the ganglionic cell layer of the retina is where the stimulus acts. Held to substantiate this is Iwama's finding that the eye's chronaxy, as as treated in the next section, is too large to be accounted for as that of the optic nerve, and that dark adaptation acts differently for electric and photic stimulation. Motokawa and Iwama (54) thought to come to closer grips with this problem by stimulating the eye with exponentially rising currents of various time constants in an effort to measure accommodation.[4] Three adaptation states were chosen: complete light adaptation to 10,000 lux, "moderate" light adaptation to a dimly illuminated white surface, and complete dark adaptation of about an hour. The presence of a threshold phosphene, as voltage was changed, was what was judged. The authors found that the value of Hill's time constant of accommodation (λ) was between 35 and 90 msec. for the completely light-adapted state, 17.5 and 25 msec. for the "moderate" condition, and 45 to 320 msec. after an hour in the dark. Bouman (4) states that this time factor decreases

with light adaptation and is about 4 msec. in light and 40 msec. during dark adaptation. These results are not easy to interpret and one can say only that neural events in the eye as revealed electrically are altered in some complex way by the presence or absence of light.

Chronaxy

More evidence of the same kind of interaction was reported by Iwama (23) in a study in which he measured the time-intensity relationship of electrically exciting the eye under three different conditions of adaptation. These were somewhat as before: 950 lux for the light condition, about three lux for the moderate, and after 40 min. in the dark for dark adaptation. Average chronaxies for these three conditions were 6.2 msec., 30.8 msec., and 13.8 msec., respectively. Such large chronaxies are not characteristic of muscle and nerve tissues and are considered by Iwama to be too large to be attributed to the optic nerve. He is of the opinion that his values result from stimulating the ganglion cells of the retina. Iwama's chronaxies are much larger than those of 1 to 3 msec. measured by the Bourguignon school (5) but are in somewhat closer agreement with Verrijp (81).

Use of Pulses

Iwama also followed the change of electrical thresholds during the course of the first 25 min. of dark adaptation. He used rectangular pulses of different durations for this. Here it was found that the thresholds rose for a few minutes after the onset of dark adaptation and then reached a constant value at times that depended on the duration of the pulses. The shorter the pulses, the earlier the leveling off occurred. With 3.9-msec. pulses, no change in threshold

---

[4] When a stimulating current is applied to a nerve tissue, excitability, after reaching a maximum value, tends to return to its initial or resting state. This effect is known as accommodation. If a current is increased very slowly, accommodation will act to prevent the threshold from ever being reached. Accommodation is complete when excitability returns to its resting value with the continued flow of a constant current. This occurs in some tissues. In others, it does not return at all. This latter is the case of no accommodation. Where there is no accommodation, a nerve will continue to give responses as long as the current flows, provided the intensity is of threshold level. The eye, incidentally, shows rather good accommodation as evidenced by the fact that a flash occurs only on the make and the break of the stimulus current.

was observed after about 10 min. of dark adaptation. Stimulation with pulses of 52.5-msec. duration delayed this effect until about 20 min. In addition, Iwama stated that the chronaxy, during dark adaptation, passed through a maximum value at about 12 min. This again shows, in a general sort of way, that there exist, somewhere in the visual mechanism, a number of interacting factors that relate photic and electric excitation.

oscillation.[5] When the reciprocals of the threshold intensities of electric test stimuli given at spaced intervals were plotted against the time after a subthreshold sensitizing current had been delivered, the result was an E or "excitation" curve. The amplitudes of the oscillations that were found depended upon the adaptation state of the eye and were noted
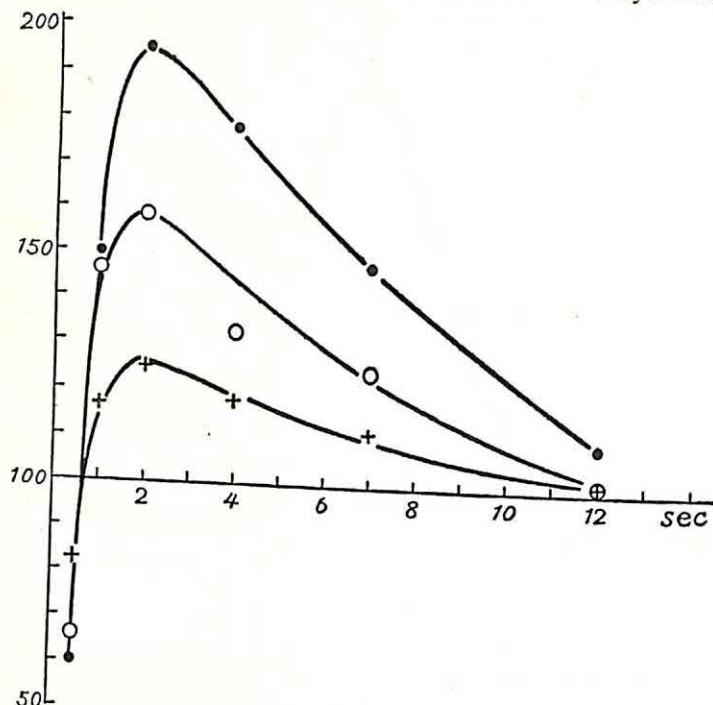

FIG. 3. E-CURVE SHOWING THE INCREASED ELECTRICAL EXCITABILITY OF THE EYE AFTER A SENSITIZING LIGHT. UPPER CURVE 70 LUX, MIDDLE CURVE 1.1 LUX, LOWER CURVE .06 LUX. ORDINATE: EXCITABILITY, PRE-ILLUMINATION VALUE TAKEN AS 100. [MOTOKAWA (39).]

Supernormal Excitability

Motokawa (38), in a detailed account of periodic excitability of nervous tissue, referred to a paper by himself and Iwama (53), in which it was demonstrated that a subnormal electric stimulus left in its wake an excitatory state in the retina that subsided like a damped

to be largest for a moderately light-adapted eye. This finding led Motokawa to suspect that photic stimulation might also leave the same kind of excitatory state (39).

The investigation of this possi-

[5] Motokawa refers again to this effect in a later paper (44) where a sensitizing current that was 80 per cent of threshold was used. Oscillations at 18 cy./sec. were observed that die away in amplitude. Certain other complex frequency relationships relating to this phenomenon were also reported.

bility consisted in dark-adapting the eye for about 20 min. Then light exposures of from .2 to 2 sec. were given, followed by tests of the electrical threshold after a variable delay. The exposures were to white light at 70, 1.1, and .06 lux, and the thresholds were determined with constant current rectangular pulses of 57-msec. duration. The data

crest in these curves was reached was independent of the intensity of the pre-illumination. It was also independent of the duration of the pre-illumination, at least at the fovea (49). If this test is made with a light stimulus, the sensitivity of the retina is, of course, always found to be lower. Motokawa again proposed that the probable locus of the elec-
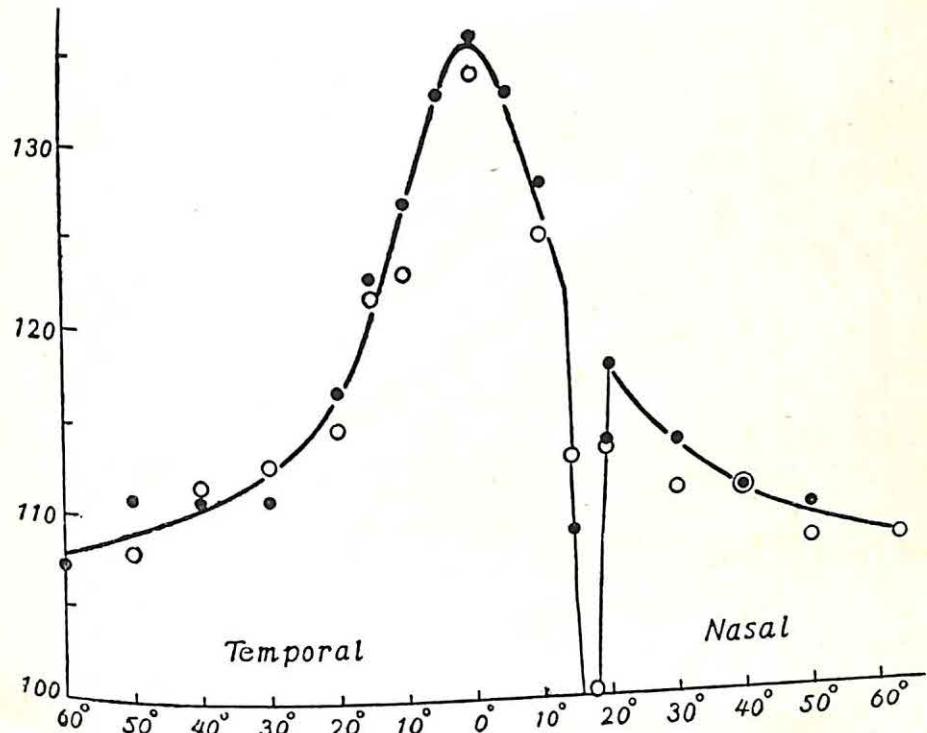

FIG. 4. EFFECT OF RETINAL POSITION ON INCREASED EXCITABILITY AFTER PRE-ILLUMINATION TO 287 LUX. ORDINATE: EXCITABILITY, PRE-ILLUMINATION VALUE TAKEN AS 100. [MOTOKAWA (39).]

are shown in Fig. 3, where the abscissa is elapsed time after pre-illumination and the ordinate is the reciprocal of the electrical threshold in percentage of what it was before illumination. The pre-illumination excitability value is taken as 100. It is seen from this curve that the eye becomes maximally excitable about 2 sec. after the cessation of white light. The time at which the

trical effect is the ganglion cells of the retina. He pointed out that the decay of an electrical sensitization process is very rapid (200 msec.), whereas the aftereffects of a photic process persist for 10 sec. or more. This difference he attributed to the involvement of photochemical reactions, and argued that if the optic pathways or higher centers were responsible, light and electricity, as

sensitizing agents, should behave in the same way.[6] He added further evidence of this by illuminating one eye only, but measuring the thresholds in both. The result showed that the enhanced excitability occurred only in the illuminated eye.

Retinal Position

It will be noted that Motokawa's method has by now taken on a dis-

for the left eye are shown in Fig. 4. Sensitivity is enhanced most at the fovea, is zero at the blind spot, and falls off markedly toward the periphery. Figure 5 shows the same measurements for electrical excitability following the use of a near-threshold intensity of pre-illumination. Curve A (solid) is the electrical data. Curve B (broken) is sensitivity to light as measured with an adaptom-
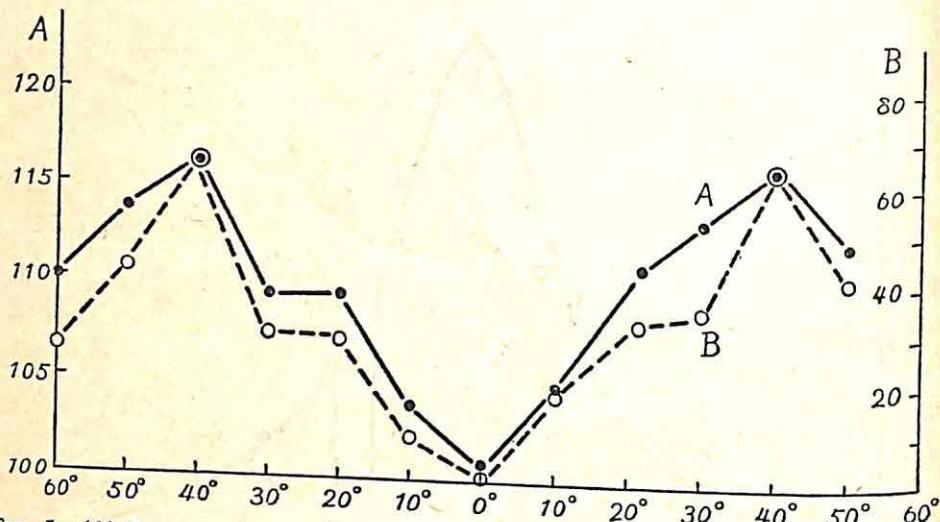

FIG. 5. (A) EFFECT OF RETINAL POSITION ON EXCITABILITY AFTER PRE-ILLUMINATION AT A NEAR-THRESHOLD INTENSITY. (B) SENSITIVITY TO LIGHT AS MEASURED WITH AN ADAPTOM-ETER. [MOTOKAWA (39).]

tinctly analytical cast. In this same paper he tested further the versatility of electrostimulation as a research tool. He chose as his measure electrical thresholds taken at the crest point on the enhancement curve; that is, measured 2 sec. after the cessation of pre-illumination. He then pre-illuminated the retina with a 2° circular spot of 287 lux at different positions on the retina. Results

[6] Another difference between electric and light stimulation is that flickering phsophenes can be experienced to frequencies of intermittent current up to 80 cy./sec. or more (31). This is far above the fusion frequency for light and appears to be due to the absence of the chemical reactions that must operate in the case of light.

eter. The curves agree remarkably well.

Intensity of Pre-Illumination

The next step was to determine the effect of the intensity of the pre-illumination on the electrical excitability of the eye (59). The test current was confined to the illuminated eye by using electrodes above the eye and at the outer angle. The test pulses were 100 msec. in length. The size of the pre-illuminating test patch was 2° for observer K.I. and 8° for K.M. It was centrally fixated. The measure ζ plotted against log I for two Os is given in Fig. 6. A word of clarification is in order here,

as this symbol will be used frequently in the discussion to follow. The measure ζ is defined as

$$\zeta = \frac{E - E_0}{E_0} \times 100,$$

where E_0 is the excitability, or reciprocal of electrical threshold, as shown by Iwama to be constant by then (23). We may now return to Fig. 6 in which the second curve is displaced to the right two log units. The dotted curve is attributed to the rod function since the 8° area covered more than the fovea in K.M.'s case.

It was pointed out that the data obtained in this experiment bore a
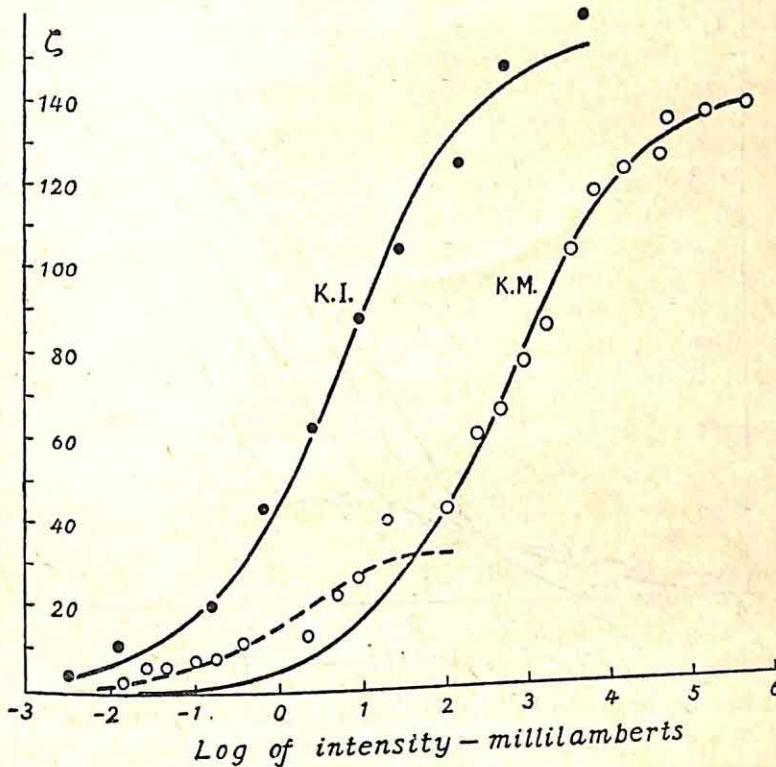
FIG. 6. EFFECT OF INTENSITY OF PRE-ILLUMINATION ON EXCITABILITY. ORDINATE: ζ IS THE PER CENT INCREASE IN ELECTRICAL EXCITABILITY DUE TO LIGHT OVER WHAT IT WAS WHEN MEASURED IN THE DARK. [MOTOKAWA AND IWAMA (59).]

determined in the dark after 20 min. of dark adaptation, and E is excitability as determined after pre-illumination has ceased. The quantity ζ, therefore, is the increase in electrical excitability due to light over the electrical excitability in the dark expressed as a percentage. Twenty minutes of dark adaptation were chosen for the base E_0, inasmuch as the threshold to electric pulses was

marked similarity to those obtained when CFF and visual acuity are plotted against light. Hecht (20) had shown that certain visual phenomena could be reasonably well described by the equation

$$KI = x^n/(a - x)^m,$$

where I is the light intensity, x the concentration of the photolytic substance in the stationary state, a is

the initial concentration of the photo-sensitive substance, K expresses the ratio of the velocity constants of the light and dark reaction, and m and n are numbers characterizing the order of the light and dark reactions, respectively. Motokawa proceeded to relate his measure ζ to this formula by substituting it for x, and the

the fovea the measures were taken. Figure 7 shows the results of measurements taken with a 2° pre-illuminating patch at the fovea, and at 5°, 20°, and 50° removed from it. Motokawa concluded that the slowly rising portion of these peripheral measures for low intensities represented rod function, and the steeper
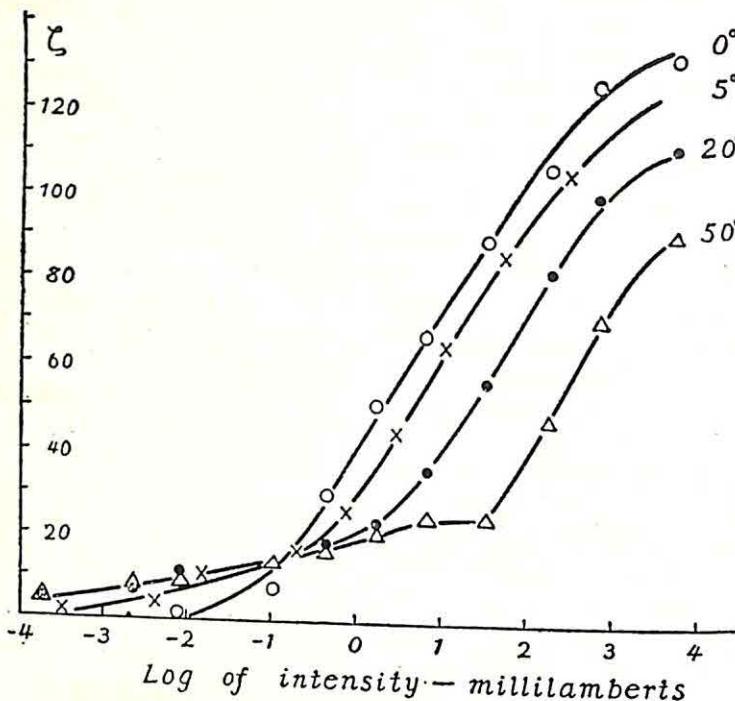
FIG. 7. ζ-LOG I CURVES FOR PRE-ILLUMINATION AT DIFFERENT RETINAL POSITIONS. [MOTOKAWA AND IWAMA (59).]

solid curves shown in Fig. 6 were computed by this means, m and n both being taken as 2 as in Hecht's case. Motokawa had suggested in the previous paper (39) that a chemical or photosensitive substance was involved that raised electrical excitability as a consequence of light action. The success of using Hecht's equation, he felt, added convincing evidence for it.

Outside the fovea, the sigmoid relationship between ζ and intensity consisted of two segments that became more distinct the farther from

portions at high intensities were due to the cones. In any case, satisfactory descriptions of this and the preceding data were obtained by substituting ζ in Hecht's equation to obtain

$$KI = \zeta^2/(a - \zeta)^2.$$

Instantaneous Thresholds

The work on the time-course of excitability following pre-illumination so far reported had not included measurements at delays shorter than

about one-quarter of a second. Since it was indicated that a short subnormal excitability period existed before the rather lengthy supernormal period, Mita, Hironaka, and Kōike (33) investigated the region between the time when pre-illumination ceased and 2 sec. Except for more elaborate timing equipment for controlling the duration of preillumination exposure and the introduction of the test stimuli, the method was as already described. A 1.5° circular field of light and 140-msec. test pulses were used. In studying the effect of intensity for light exposures of 1/250 sec., it was first noted that at all intensities the value of ζ was negative (subnormal excitability) for about .75 sec. It then rose to a maximum at about 2 sec., as had previously been shown. In addition, the finding was verified that the maximum value reached by ζ increased with the intensity of preillumination. During the subnormal period, however, the reverse occurred. The lowest values of ζ were reached when the highest intensities were used. Another aspect of this study showed what happened when intensity was held constant and the exposure time was varied. The highest supernormal values of ζ were reached with long exposures. The results for the subnormal period are confusing, although there is an indication that the length of the subnormal period may be inversely related to the length of the light exposure.

"Empfindungszeit"

Bearing upon another aspect of the general problem of the excitability of the retina is a rather uncertain paper by Hironaka (22). He pointed out that the problem of where the electric stimulus acts in the eye is still unsolved and suggested the comparison of the *Empfindungszeit* for a photic and electrically induced sensation as a means of settling the matter. The *Empfindungszeit* (EZ) was proposed by Fröhlich as a term for the time required by a stimulus to evoke a sensation (10). The photic EZ had been measured by Mita, Hironaka, and Kōike (34) for different conditions of adaptation and different retinal positions in a previous contribution. Hironaka employed the same method for the phosphenes and reported that EZ's for light at the fovea varied from 67.5 to 185 msec. depending on the light intensity, and for phosphenes in the dark-adapted eye from 68.5 to 176.8 msec. depending on the current. These agree well. For EZ's taken 50° out on the periphery, however, the values were about 358 msec. for light. Since this is about twice the EZ for the phosphenes, Hironaka related the foveal light EZ with the effect of electric stimulation and assumed that the electrical excitability and speed of reaction of the cones were higher than that of the rods.

It is not completely clear that this is the case, although there is evidence that the scotopic process is slower than the photopic. The fusion frequency for light is lower in the peripheral retina than in the central region, and dark adaptation proceeds slower in the rods than in the cones. In addition, Motokawa and Ebe (49) have presented data to show that the rod process, as determined by the ζ-time method, is the slowest in the retina. On the other hand, work by Bourguignon and Déjean (5) and Kleitman and Piéron (25) shows the speed of response of the cones to be less than that of the rods, whether measured by light or electric stimuli. Whatever the case, the problem of measuring the speed with which sensations and perceptions are
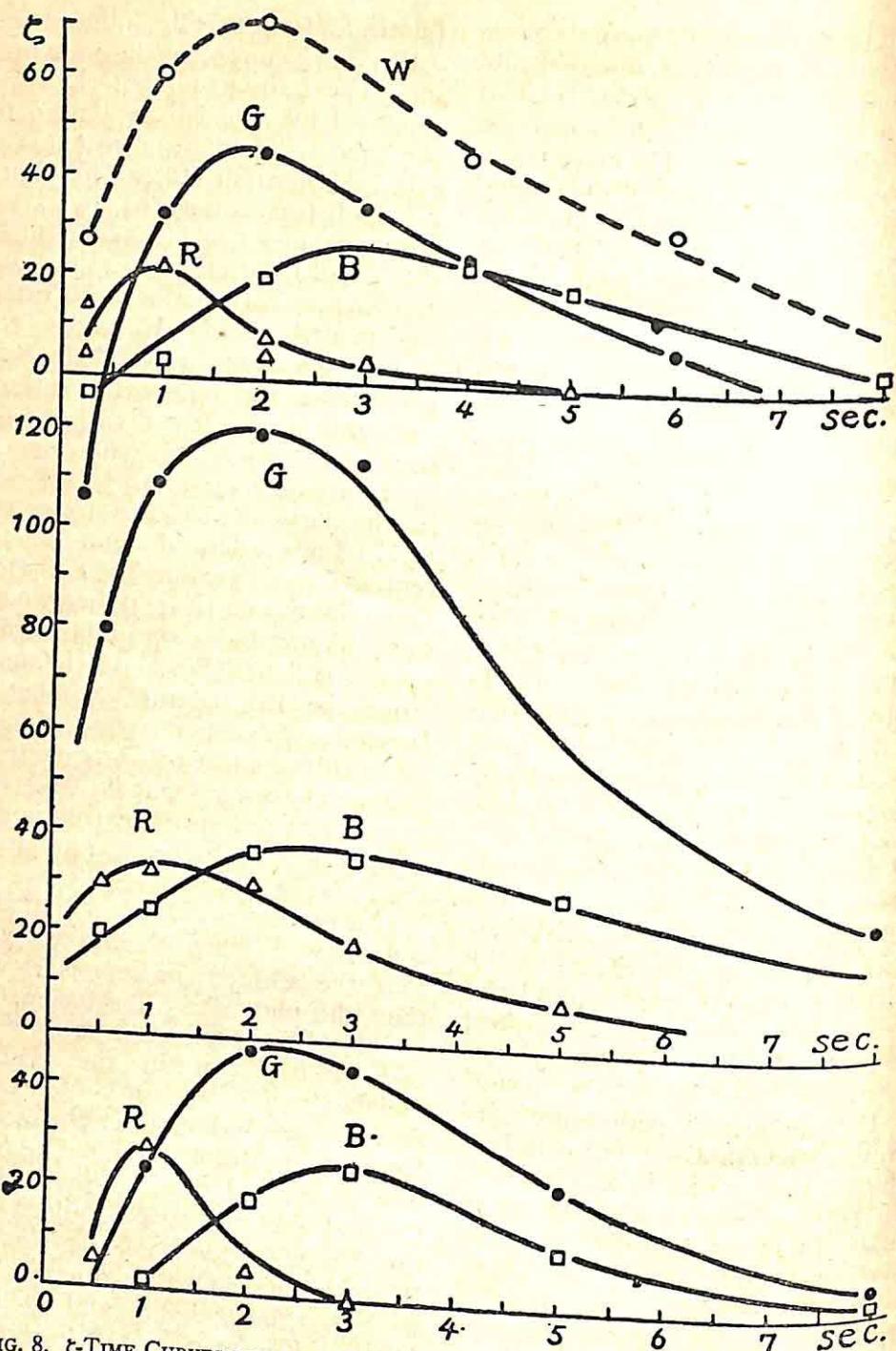
FIG. 8. ʒ-TIME CURVES FOR THE NORMAL EYE WITH RED, GREEN AND BLUE LIGHTS OB-
TAINED BY FILTERS. THREE DIFFERENT SUBJECTS. [MOTOKAWA (40).]

aroused has always been a knotty
one and it is not surprising that
agreement has not yet been reached.

II. *Color Discrimination*
The results of the physiological
aftereffects of brief illumination with

white light suggested, quite naturally, that a similar examination be conducted with colored lights. The time-course curve of ζ that showed the retina to be at maximum excitability 2 sec. after the termination of neutral pre-illumination might well be of different shape with selective stimuli. The method used to investigate this aspect of the problem was the same, in the important details, as has been described for the white light studies.

Hue and the Time-Course of Electrical Excitability

In the first study of this kind reported by Motokawa (40), both filtered light and monochromatic light from a dispersion spectrum were used. The pre-illumination was a spot of 2° exposed for .2 sec. and the electrical testing was done with 100-msec. pulses. Samples of filtered light data for three different subjects are shown in Fig. 8. The filters used were red (620–740 mμ), green (490–550 mμ), and blue (410–490 mμ). It is to be noted that the sets of curves differ markedly in shape. This matter of striking individual differences suggests color anomalies in vision, a problem to be taken up in the next section. But of primary interest here is the fact that the maxima of the red, green, and blue curves are reached at different times; namely, at about 1, 2, and 3 sec., respectively. The white curve, shown in the upper part of the figure, corresponds to the green maximum except for amplitude. When Motokawa compared the foveal and peripheral regions by pre-illuminating with filtered lights, he found that the ζ-time curves taken at 40° out on the periphery were very nearly the same for all hues and that in each case the maximum occurred at 2 sec. The relevance here for color theory is obvious.

Effect of Intensity

In this same paper was shown the effect on ζ of reducing the intensity of spectral lights of 650, 585, 530, and 470 mμ. The published curves indicate that the maximum value of ζ for any given wave length underwent no shift in time as the intensity was reduced. It seemed to be a property of the wave length alone. The yellow light at 585 mμ afforded a slightly later maximum than the red and was located between the red and green at about 1.25 sec. to 1.50 sec.

Motokawa emphasized in this paper that time in these curves concerns the rate of the recovery process of the retinal physiological mechanism after illumination: the longer the wave length, the more rapid the recovery. Intensity concerns the magnitude of physiological process but not its rate. Time and intensity, therefore, are shown by the data, to be independent.

Wave Length and the Maxima of the ζ-Time Curves

The times after pre-illumination at which the maxima of the ζ measure were reached were investigated by Motokawa (42) for spectral light of different wave lengths. It was found that the time, τ, at which the maxima occurred, varies with wave length, λ, as shown in the upper curve of Fig. 9. Time, τ, varies between something over 3 sec. for light of 400 mμ and about 1 sec. for light of 700 mμ. The curve connecting these two extremes is not smooth and τ changes rapidly with respect to λ in several places. The next curve shown in Fig. 9 is for $\Delta\lambda$ and was obtained by conventional methods of measuring hue discrimination. The data are for Motokawa's eye and show minima at 460, 490, 585, and 630 mμ. Final-

ly, the curve $\Delta\lambda/\Delta\tau$ was obtained by plotting the decrease of wave length $\Delta\lambda$ corresponding to the reciprocals of the slope of the τ curve against λ. This curve corresponds reasonably well to the hue discrimination function and is taken to be further evidence that a basic physiological process is being tapped by the electrical

color vision was the same as has been described for preceding studies on the determination of the ζ-time curves. Figure 10 from Motokawa (43) is the excitability curve of a deuteranomalous subject measured after pre-illuminating the center of the eye with white light. Three maxima appear at times correspond-
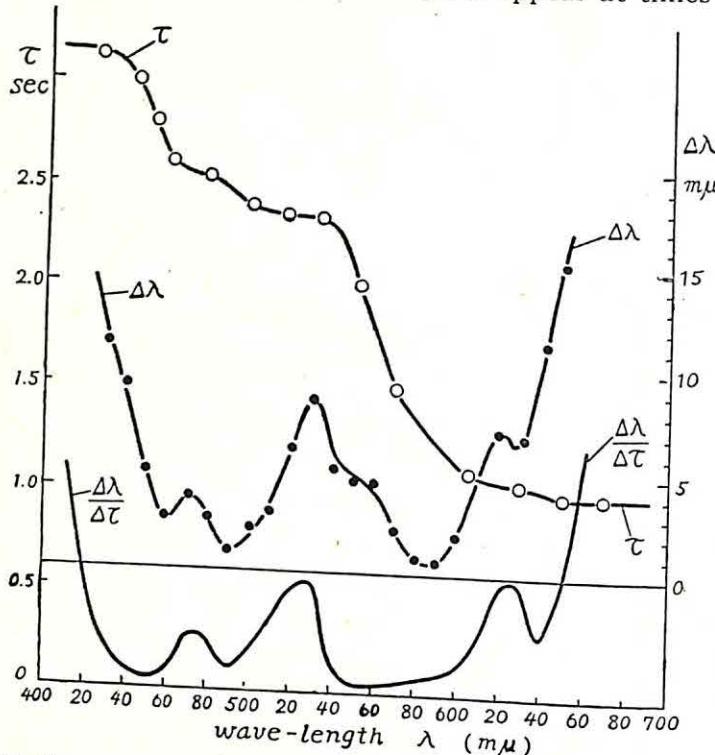

FIG. 9. UPPER CURVE: THE TIME (τ) AT WHICH THE MAXIMUM FOR ζ OCCURS FOR EACH WAVE LENGTH (λ). MIDDLE CURVE: CONVENTIONAL HUE-DISCRIMINATION. LOWER CURVE: THE CHANGE IN WAVE LENGTH ($\Delta\lambda$) CORRESPONDING TO THE RECIPROCALS OF THE SLOPE OF τ FOR EACH WAVE LENGTH (λ). [MOTOKAWA (42).]

method. Once more it was asserted that this mechanism is peripheral and lies in the ganglion cell layer of the retina.

Color Deficiency Studies

The foregoing success in isolating distinctive patterns of electrical thresholds for selective illumination of the retina next led to the important question: What can be learned from cases of color blindness? The method used to examine persons of deficient

ing to those obtained by applying red, green, and blue lights separately to a normal eye. A normal trichromat, it will be remembered, manifests only one maximum to white light. Motokawa noted that for the deuteranomalous case the more weakly excitable G curve dropped out rapidly as intensity was reduced. He felt that he had here basic physiological evidence of the three independent processes assumed by the three-components theory of color

vision. More evidence of this same kind is given by data presented in later papers by Motokawa (**35, 36**), Motokawa and Suzuki (**65**), and Ebe, Isobe, and Motokawa (**9**), where the method was extended to measurements on deuteranopes and pro- sively displaced forward in time as wave length is shortened. This fact has been discussed in the preceding section (**42**). Motokawa assumes here that the continuous ζ-time curve of white light is a composite of the individual curves that result from
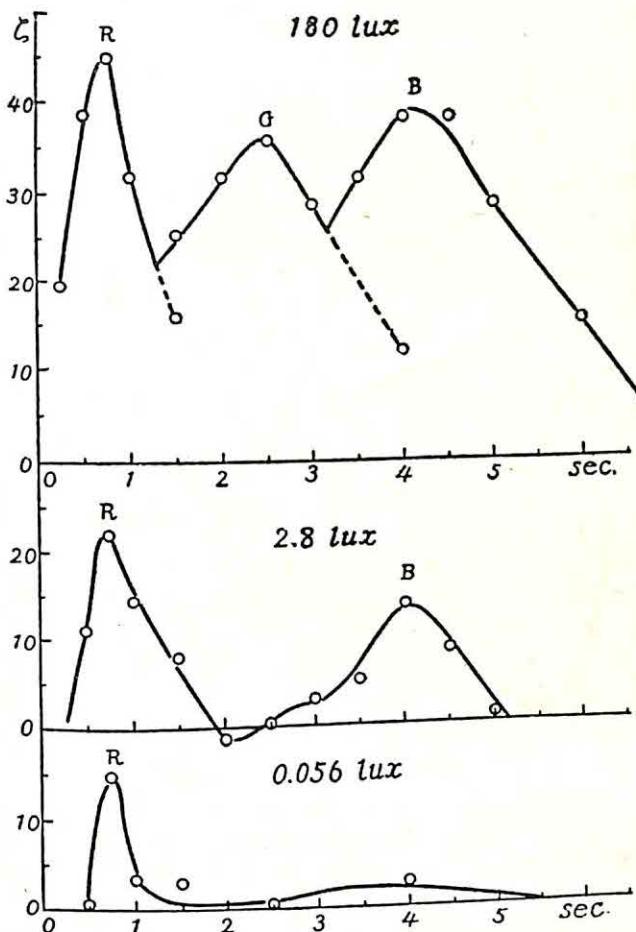
FIG. 10. ζ-TIME CURVES FOR THE DEUTERANOMALOUS EYE OBTAINED WITH WHITE LIGHT. [MOTOKAWA (**43**).]

tanopes, as well as normal eyes measured when pre-illumination was applied outside the fovea. Figure 11 from Motokawa (**35**) shows a further analysis of what happens in the deuteranomalous eye and compares it with normal color vision. For the normal eye, light of different wave lengths produces maxima in the ζ-time curve that are progres- selective stimulation. For the deuteranomalous case, however, the resulting family of curves reveals excitability maxima that at all times reflect the weakness of the underlying physiological G process. For white light these sensitivities fuse into a composite ζ-time curve with three maxima.

To show the effect of retinal posi-

tion, Motokawa stimulated the normal eye extrafoveally at 5°, 10°, 15°, 25°, 35°, 45° and 50°. The pre-illuminating stimulus was a 2° patch of white light. The results are shown in Fig. 12. At 5° and 10° a curve of three maxima is seen. This is

a single maximum at 2 sec. The nature of this residual excitation curve is thought of as a rod mechanism.

Physiological Sensation Curves

In a further attempt to isolate the nature of the basic physiological
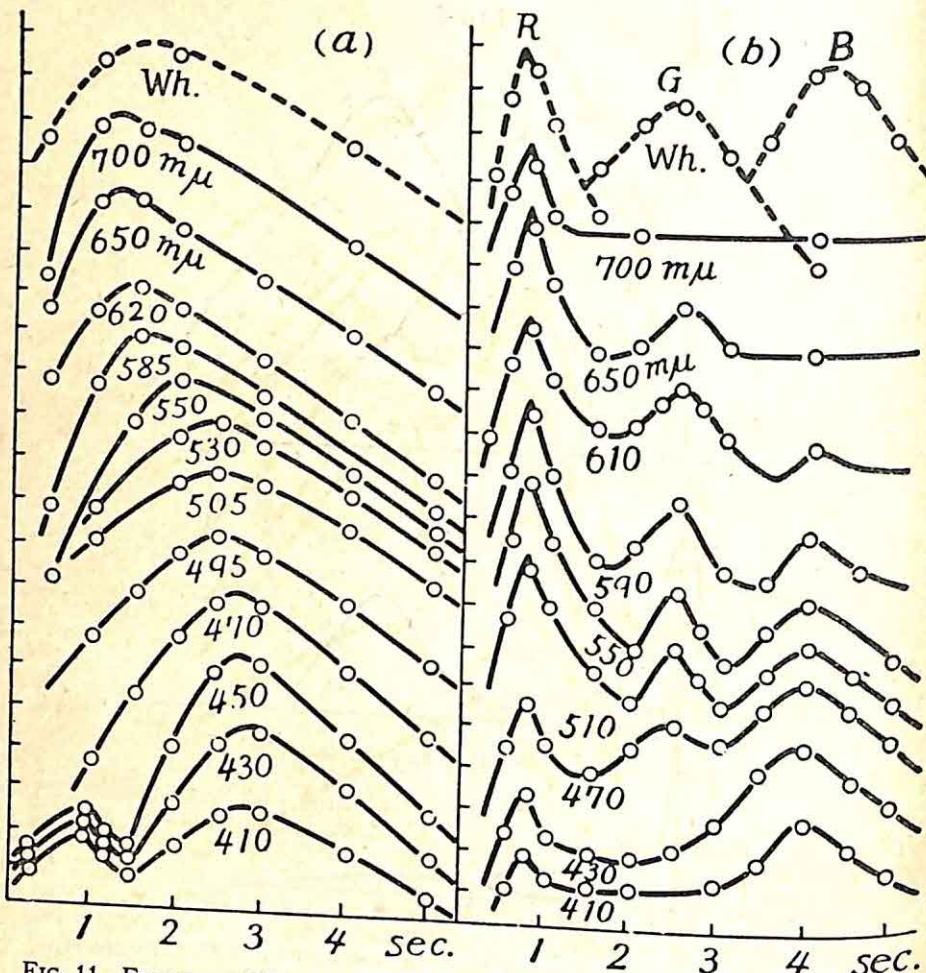
FIG. 11. EFFECT OF WAVE LENGTH ON THE ζ-TIME CURVES OF NORMAL (a) AND DEUTERANOMALOUS (b) EYES. [MOTOKAWA (35).]

taken to be indicative of deuteranomaly in the parafovea of the normal eye. Further out on the periphery, the R and G maxima both drop out, leaving those corresponding to the Y and B. Up to this point the electrical findings correlate with what is found in ordinary perimetric studies. At 50° a ζ-time curve results showing

processes in vision by electrostimulation, Motokawa (35) performed a very interesting experiment. He had previously determined the times after pre-illumination at which the maximum of ζ occurred as a function of wave length. Now he explored wave length, using the three times at which the maximum to red, yellow, green,

and blue light occurred. These were taken as 1, 1.5, 2, and 3 sec., respectively. The upper set of curves in Fig. 13 is due to illuminating the fovea of the normal eye and shows how ζ varies with wave length when measured at these times. Motokawa

and 4 sec. for B. The results for each wave length may be seen by comparing Fig. 13 and 14 for the trichromat and deuteranomalous cases, respectively.

First, it is seen in Fig. 13 that for the normal subject the B process
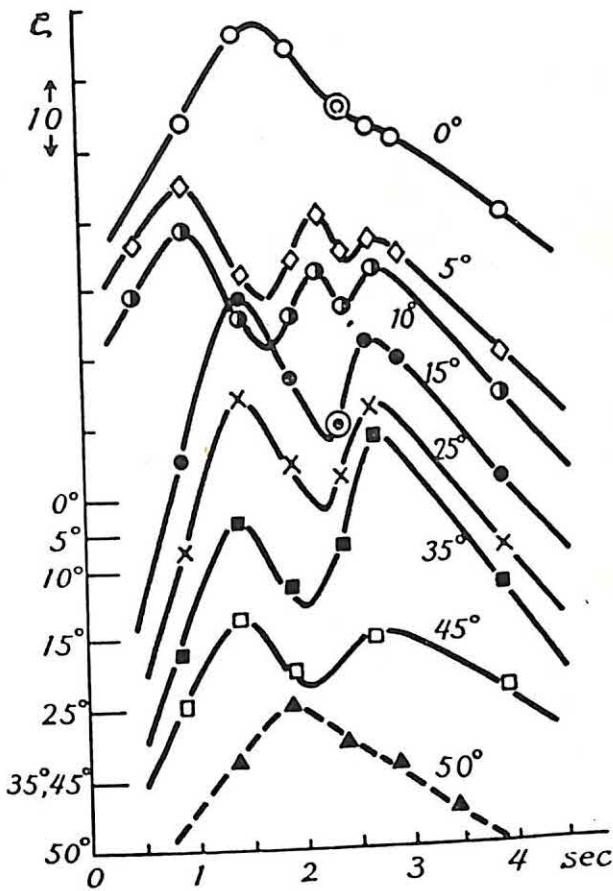
FIG. 12. EFFECT OF RETINAL POSITION ON THE ζ-TIME CURVE OF THE NORMAL EYE. [MOTOKAWA (35).]

calls these functions physiological sensation curves and points out their striking similarity to those classically obtained by color mixture. He then proceeded to use this method of analysis on both normal and deuteranomalous eyes at different positions on the retina: 0° (fovea), 15°, 25°, and 35° (36). For the color-deficient case the time values were: .75 sec. for R, 1.5 sec. for Y, 2.5 sec. for G,

holds up well until stimulation is applied in the extreme periphery. The R and G processes, meanwhile, diminish in excitability. The broken line with the filled circles among the foveal curves of Fig. 13 represents data taken with a time delay of 1.5 sec. Here the Y process curve shows two humps, one each at R and G. This was interpreted by Motokawa to mean that the physiological
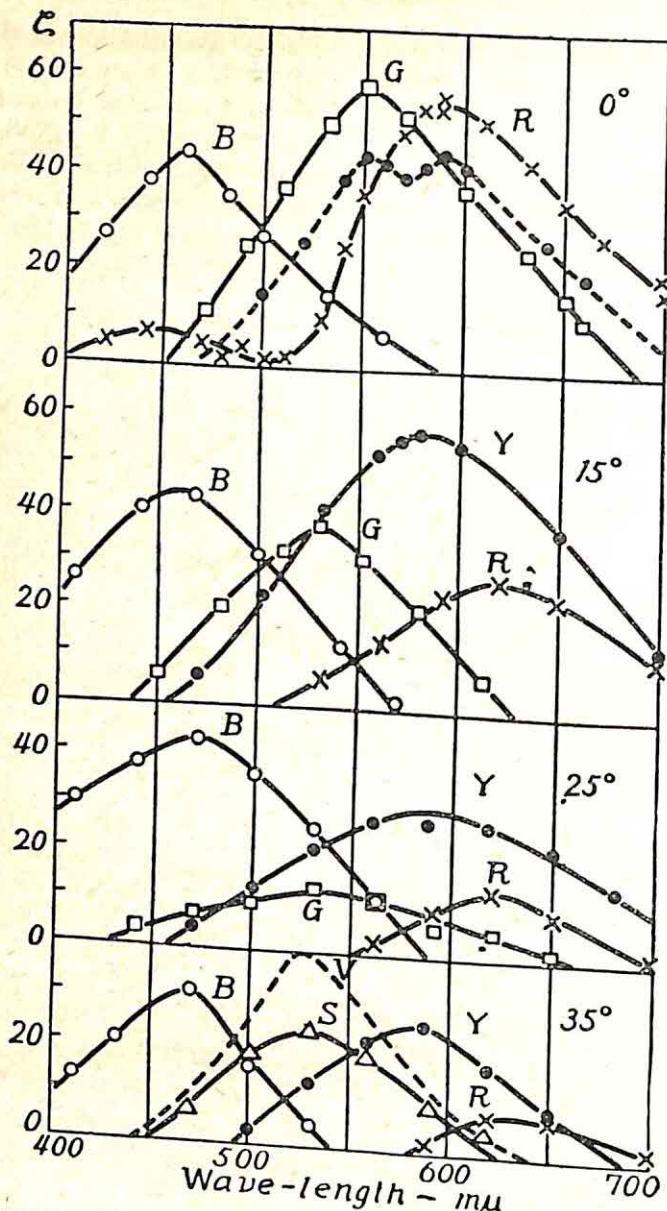
FIG. 13. EFFECT OF RETINAL POSITION ON THE PHYSIOLOGICAL SENSATION CURVES FOR THE NORMAL EYE. [MOTOKAWA (36).]

method reveals no independent Y process in the fovea. Outside of the fovea, however, the Y process at 15°, 25°, and 35° shows up to be stronger than the R and G. In the fovea, the R and G have maxima somewhat different from what appears parafoveally, and the R has a slight rise in the short wave lengths at about 440

mμ. Another point to note is that at 35° the middle of the three processes is elevated over what it was at 25°. This, Motokawa marked S (scotopic) to distinguish it from the G process. The broken curve, V, in the bottom section of Fig. 13 is a scotopic curve determined with a flicker photometer. Since both S and V are

seen to cover the same spectral range with maxima at the same position, they are interpreted as a reflection of rod activity.

The deuteranomalous data in Fig. 14 need little comment. The Y process is shown as absent from the fovea, but appears in the periphery.

ζ-time curves showed the protanope to have a depressed excitability at about 1 sec. and the deuteranope at about 2 sec. In the physiological sensation curves, the protanope lacked the R process curve and the deuteranope the G curve. Data on the physiological sensation curves
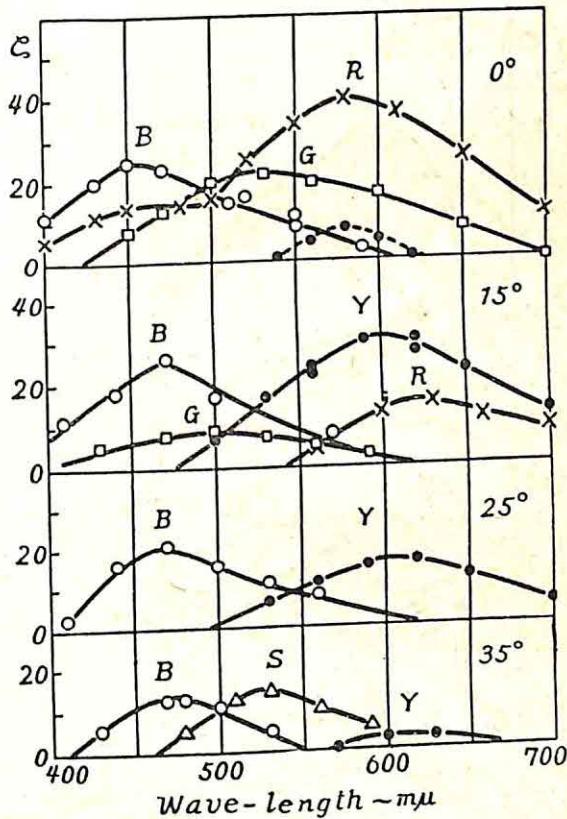

Fig. 14. Effect of Retinal Position on the Physiological Sensation Curves for the Deuteranomalous Eye. [Motokawa (36).]

The weak G process diminishes rapidly and drops out at 25°. The S curve again turns up at the extreme periphery.

Dichromats

Data on dichromats were presented in several papers (9, 36, 65). In general, alterations in the ζ-time and physiological sensation curves were as would now be expected. The

for dichromats were very clearly in agreement with prediction. In the ζ-time data as presented in papers by Motokawa (36) and Motokawa and Suzuki (65), there appears to be contradiction in the way the protanope behaves. More helpful are the data given by Ebe, Isobe, and Motokawa (9), where it is made quite evident that excitability in the protanope is shifted to the longer times

corresponding to the shorter wave lengths.

Motokawa and Suzuki (65) reported that if, in the case of the deuteranomalous S, the intensity or size of the pre-illuminating stimulus was reduced, the weak elevation in the ζ-time curves disappeared and the

Inhibitory Action of Colored Light

Motokawa opened up a new area for his electrovisual investigation by the accidental discovery of what he calls selective inhibition (47). During some experiments on summation (46) he happened to be working in a room that was not completely
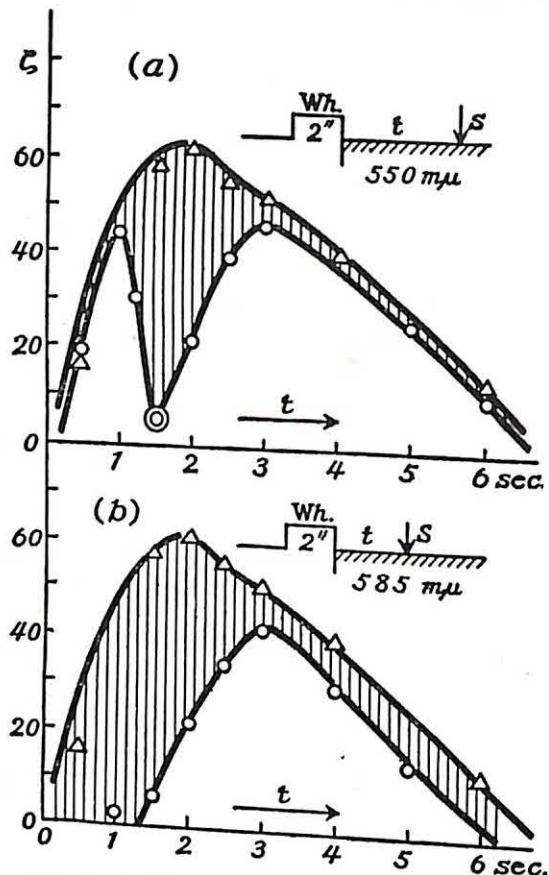
FIG. 15. ALTERATION OF THE ζ-TIME CURVE OF WHITE LIGHT BY INHIBITORY COLORED LIGHTS. THE UPPER CURVES ARE THE NORMAL ζ-TIME RELATIONS FOR THE WHITE LIGHT. THE LOWER CURVES ARE DUE TO LIGHT OF 550 AND 585 mμ. THE SHADED AREA IS THE AMOUNT OF INHIBITION. [MOTOKAWA (47).]

result could not be distinguished from that of a deuteranope. In this paper they also presented further data on the ζ-λ relationship for dichromats and anomalous trichromats.

It is worth mentioning, in passing, that the electrostimulus method is claimed to be more sensitive than conventional tests in detecting the presence of color deficiency (36).

dark but was faintly illuminated by green light. The effect of this changed condition was to deform the ζ-time curve in this way: In the normal eye with white light the presence of the green illumination caused the curve to exhibit two maxima instead of one and thus to resemble data obtained from a deuteranope. Two hypotheses to explain the disap-

pearance of part of the curve were entertained. First, that the effect was simply selective adaptation. This was rejected when it was found that the green light had to be present during the time course over which excitability was measured. It could not be removed with the termination of pre-illumination. The second hypothesis was that the green light had a selective inhibitory influence on the development of the G process.

An example of these kinds of data obtained at the fovea of the normal eye is shown in Fig. 15 where in the upper curve light of 550 mμ is seen to reduce greatly the G process and leave the R and B processes intact. The shaded area represents the inhibitory effect. In the lower curve light of 585 mμ eliminates the R and G processes and leaves the B process. Motokawa's theory was that if it is assumed that white light arouses the R, G, and B components in some def-

inite proportions and that the inhibitory light withdraws some of the excitation process from particular components, the excitability curve would be the result of what is left. This process apparently operates on a complementary basis.

Figure 16 shows how an analysis by inhibitory lights applies to yellow pre-illumination of 589 mμ in which the maximum for ζ in the normal fovea occurs at about 1.5 sec. as shown in curve A. The effect of light of 550 and 530 mμ is to inhibit the G component of yellow light, but not the R. Light at 430 mμ takes out some of the R and B, and leaves the G intact. A mixture of 430 mμ and 680 mμ eliminates the R completely and leaves the G. Motokawa felt that this again demonstrated the dual nature of the effect of yellow light in the fovea.

For the parafovea, where he reported having found a separate Y process (36) in both normal and deu-

FIG. 16. ANALYSIS OF LIGHT OF 589 mμ BY SELECTIVE INHIBITION. [MOTOKAWA (47).]

teranomalous eyes, he concluded on the basis of a similar analysis using selective inhibition that the Y process is unitary and not composed of R and G. However, it is not clear from the data presented in this paper that this is the case.

The next report on this matter was ference in ζ-units between a normally determined excitation value and that obtained when the inhibitory light was used. An example will help here. If Motokawa wished to measure the inhibitory effect of light of any wave length on the R process, the ζ-value taken 1 sec. after
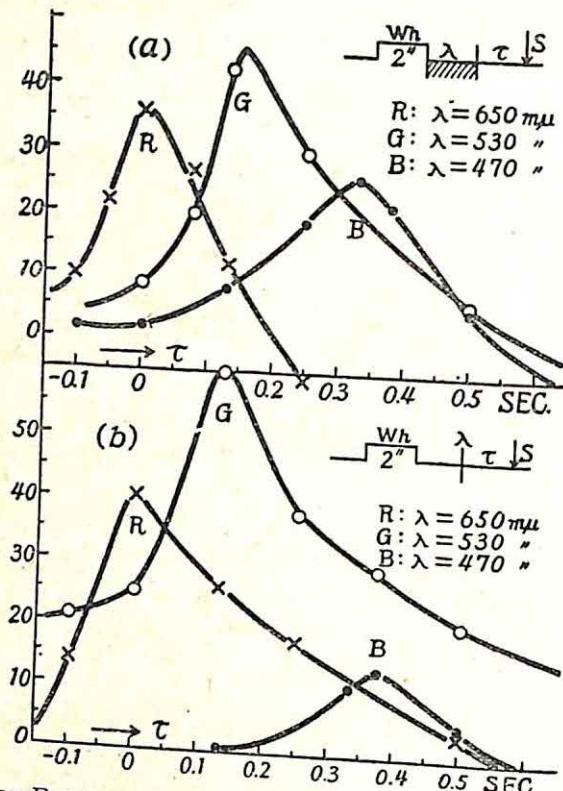
FIG. 17. INHIBITORY EFFECTIVENESS OF COLORED LIGHT ON THE R, G, AND B PROCESSES OF THE NORMAL EYE. SEE TEXT. ABSCISSA: THE INTERVAL BETWEEN THE INHIBITORY LIGHT PULSE AND THE ELECTRIC STIMULUS. [MOTOKAWA AND SUZUKI (67).]

by Motokawa and Suzuki (67) and dealt with the problem of where, in time, to introduce the inhibitory light for greatest effectiveness. The first thing the authors did was to measure inhibitory effectiveness, as related to the R, G, and B processes, by presenting a flash of inhibitory light at varying intervals before and after the electric stimulus was given. The measure ζ was again used and inhibition was expressed as the dif-

stimulation with white light would first be found. This would be maximum for a normal eye. Then an identical measurement of ζ would be obtained with pre-illumination ceasing at a moment .5 sec., say, prior to the electric stimulus, which was always delivered at precisely the 1-sec. mark. The difference between these two values of ζ would be the inhibitory effectiveness. To study the inhibitory effect on G and B

processes, the same thing would be done, except that 2 and 3 sec. would have to be used. The interval between the inhibitory flash and the electric stimulus was called the inhibition-electric stimulus (I-S) interval. It was considered positive when the flash preceded the current and negative when it followed. Figure 17 shows the effectiveness of

Stimulation of the periphery, 15° from the fovea, revealed a fourth I-S interval as measured from a base 1.5 sec. after pre-illumination. This is the locus of the ζ-time maximum of the Y process. This I-S interval was at .07 sec. and was noted to lie between the R and G intervals.

At the extreme periphery, 40° from the fovea, another optimum I-S

TABLE 1

CHARACTERISTICS OF THE BASIC RETINAL PROCESSES OF NORMAL MAN AS REVEALED
BY ELECTRICAL STIMULATION OF THE EYE
[FROM MOTOKAWA AND SUZUKI (67)]

Kind of Retinal Process	True Crest Time in Seconds	Optimum I-S Interval in in Seconds	Maximum of Spectral Locus
Photopic Processes			
R	1.0	0.00	590 mμ in fovea 620 mμ in periphery
Y	1.5	0.07	585 mμ
G	2 (2–2.25)	0.15	550 mμ in fovea 530 mμ in periphery
B	3 (2.75–3)	0.35	460 mμ
Scotopic Process			
S	2.0*	0.05	520 mμ

* This excitation time is properly 4.5 sec., according to more recent measures (49, 51). See below.

light of 650, 530, and 470 mμ on the R, G, and B processes when the fovea of a normal eye is illuminated. The upper curves are the result of presenting the inhibitory light immediately after pre-illumination and turning it off at fixed times before the electric pulse. The lower curves are the effect of a 1/20-sec. flash. Both sets agree to the extent that the optimum I-S intervals turn out to be zero, .13, and .37 sec. for R, G, and B, respectively.

interval appeared at .05 sec. Since this new interval could not be demonstrated to be affected by the wave length of the inhibitory light, it was suggested as being due to scotopic or rod action.

Table 1 summarizes the characteristics of the basic retinal processes in the normal eye as shown by Motokawa's method using a 2° stimulus spot. The authors felt that the time relations in photopic vision stood together quite well but that the

scotopic process was a thing apart. Classical attempts to determine whether photopic or scotopic reactions are the faster were claimed to be irrelevant. The speeds of these reactions were shown to overlap, depending on what aspect of the photopic process was measured. Nevertheless, Motokawa and his associates proceeded to examine the nature of the scotopic process further in several recent papers.

Motokawa and Ebe (49) and Motokawa, Ebe, Arakawa, and Oikawa (51) reported, first, that while the ζ-time curves at the fovea depended little upon the duration of pre-illumination, those at the periphery had their maxima shortened by increasing durations. But the maxima remained the same if measurements were dated back to the *onset* of pre-illumination. This finding was correlated with "on"-type elements associated with the rods. Second, the true maximum time for the photopic process was amended to 4.5 sec., or the slowest in the retina (cf. Table 1). Third, when ζ was measured at 4.5 sec. for flashes of spectral light delivered to a retinal area 20° from the fovea, the resulting ζ-λ curve coincided almost exactly with the scotopic visibility curve.

In another paper Motokawa and Suzuki (66) continued the attack on the quantitative aspects of inhibitory lights. What they considered was the way intensity and wave length of inhibitory stimuli affected the amount of inhibitory effect. The experiments were done with the optimum I-S intervals for the R, Y, G, and B processes as described above. The general results indicated, first, that inhibition increased with an increase in the intensity of the inhibitory light. This was true for all of the color processes.

Second, inhibition was related to wave length more or less the way ζ is related to wave length. Therefore, the inhibition distribution curves closely resembled the physiological sensation curves shown in Fig. 13. Indeed, the maxima of these curves were claimed to correspond within the limits of experimental error.

Microstimulation Studies

Two studies have been carried out by Motokawa, *et al.* (50, 52) using the microstimulation technique of Hartridge (19) to obtain small areas of pre-illumination. The work on electrical excitability reported up to this point had been done with test patches of 2°. Now a pre-illumination area of 2' of arc was employed in, or near, the center of the fovea and at 20° from it. It was found that in the fovea the R, Y, G, and B processes existed as revealed by the ζ-time and ζ-λ curves. At the very center, however, no measurable Y response and only a weak B were observed. Near the center, or out to about 1°, the R and G process responses predominated, and beyond 2° the Y and B processes took over. The maxima of the three physiological sensation curves obtained with the 2° patch had been 585, 550, and 460 mμ for R, G, and B, respectively. For the 2' patch these become four in number at about 610, 575, 520, and 460 mμ for R, Y, G, and B, respectively. The three-components maxima were observed to correspond well both with Granit's maxima of 600, 530, and 460 mμ for the spectral sensitivity of modulators in the cat's eye, and with the general maxima in the sensation curves found by numerous investigators for color matching.

Work at 20° from the fovea confirmed the strong Y and B process

excitability and the weak R and G. The conclusion reached from these studies was that they confirmed and extended Hartridge's observations on the nature of the foveal receptor mechanism.

Evidence from the Retinae of Frogs

Since this review is intended to treat the results of electrical analysis of the human eye, no attempt is made here to do other than mention the work that Motokawa and his associates have done on frogs and toads. It should be noted, however, that seven papers have been submitted to show that much the same relationships found in human eyes can also be demonstrated in frogs (35, 55, 56, 57, 61, 63, 79). The method was essentially the same as has been described, except that the indicator of excitation now is the action potential taken from the optic nerve. Barring the difference that the action of the amphibian eye was sluggish, the ζ-time curves showed about the same characteristics as the human curves. The white, R, G, and B elevations of the ζ-time curves could be clearly seen in the data, and the dependence of ζ on intensity was evident. The phenomena of selective inhibition were demonstrated in a manner analogous to that found in the human eye. Indeed, even color contrast, to be described for human subjects in the next section, was observable in the responses of the frog's retina.

The existence of these and other phenomena in the nonhuman eye was emphasized by Motokawa as indicating anew the basic physiological nature of his electrostimulation method. It is of especial interest that work by Motokawa, Iwama, and Tukahara (63) and Tukahara (79) on modulators and dominators

gave results that correspond well with Granit's studies.

III. *Summation, Contrast and Optical Illusions*

The generality of using an electrically aroused phosphene in investigating the effect of photic ex-

TABLE 2

SUMMATION EFFECTS IN VALUES OF ζ FOR DIFFERENT DEGREES OF SEPARATION BETWEEN STIMULI
[MOTOKAWA (41)]

Angular Separation in Degrees	Single Spot	Five Spots	Summation Effect
0	54.5	77.0	22.5
1	54.5	75.5	21.0
2	54.5	68.0	13.5
3	54.5	64.0	9.5
4	54.5	58.5	4.0
6	54.5	54.5	0.0

citation processes was shown next to extend to such phenomena as summation, simultaneous and successive induction, and numerous optical illusions.

Spatial Summation

The first paper on summation was by Motokawa (41), who studied the effect of using, singly and together, arrangements of variously separated test patches. Pre-illumination with a stimulus pattern of four spots regularly arranged around a center spot was compared with a single spot for the effect on ζ. Each spot was a circle of 1° in diameter, and the angular separation between the four outside circles and the center patch was varied. The intensity of the test patterns was 1000 lux and the exposure was .13 sec. Table 2 shows the summation effect due to the use of

five spots over one spot. Summation is seen to disappear when the separation in the multiple-spot figure is about 5°. That the amount of summation depends upon intensity was shown by separate measurements using 5200, 280, and 180 lux. These intensities gave limiting values of separation of 13.6°, 7°, and 3.5°, respectively. It was pointed out that Granit (14) found summation in a four-spot experiment, using flicker, with a separation of 2° and an intensity of 100 lux.

The effect of retinal position was measured with one or both halves of a circular patch. When both halves were used, they were separated by 2°. The results appear in Table 3. The

TABLE 3

SUMMATION IN VALUES OF ζ AT DIFFERENT POSITIONS ON THE RETINA [MOTOKAWA (41)]

Degrees from Center	Single Patch	Double Patch	Summation Effect
0	63.5	69.0	5.5
20	50.0	55.5	5.5
40	42.0	48.0	6.0

intensity was 1000 lux. Here the ζ-values are seen to decrease, but the summation remains the same, indicating a greater percentage increase at the periphery than at the center of the visual field. When the intensity was reduced to 8 lux, however, the summation effect could no longer be seen beyond 10° distant from the fovea.

ζ and the Size of the Area Stimulated

This relationship was studied at 1000 lux for the fovea and at 20° and 40° in the periphery. The exposure time was .13 sec. The results for a variation in spot diameter of

one log unit are curves of the same general shape as the ζ-log I curves shown in Fig. 6.

Summation of Colored Light

Motokawa found scant evidence of this kind of summation in an experiment in which he had a split-field, circular patch of 2° diameter that could be illuminated with either one or two colors. Filters were used to obtain red (620–740 mμ), green (490–550 mμ), and blue (410–490 mμ) illumination. Stimulation was confined to the rod-free area. The effect of homogeneous illumination of the whole circle was to increase markedly the ζ-value. This doubling of the area caused ζ to rise from 42 to 75 for red, from 43 to 71 for green, and from 45 to 69 for blue. When the patch was illuminated with pairs of lights of different hues, it was found that practically no interaction existed when the colors were spectrally separated as in the red and blue. Some occurred to raise the ζ-value when the pairs lay closer together. This is explained by Motokawa in terms of the three-component theory on the basis that individual lights will stimulate more than one color process when they are near each other in wave length, but not when they are widely separated.

In another experiment Motokawa (46) looked into the matter of summation by direct color mixing where the stimulus arrangement allowed no separation between the areas where light fell on the retina. He found that a mixture of yellow (580 mμ) and blue (450 mμ) produced no significantly larger values of ζ than was true when the two lights were taken separately. This was also true when wave lengths of 650 and 470 mμ were mixed. He concluded that the law of additivity normally found for color mixing did not hold for the physiological meas-

ure of excitation as determined by his method.

The next general problem that Motokawa treated was that of brightness and color contrast (37). The method again was a simple extension of his general procedure. What he did in this case was to measure ζ by introducing the electric test stimu-

lus, as indicated in Fig. 18, where the results are also shown. ζ is raised, when the black stimulus precedes the white (upper curve), over what it is when white alone is used (lower curve —open circles). To rule out the effect of adaptation, the screen was darkened for 2 sec. before the white patch was given. This, as shown in
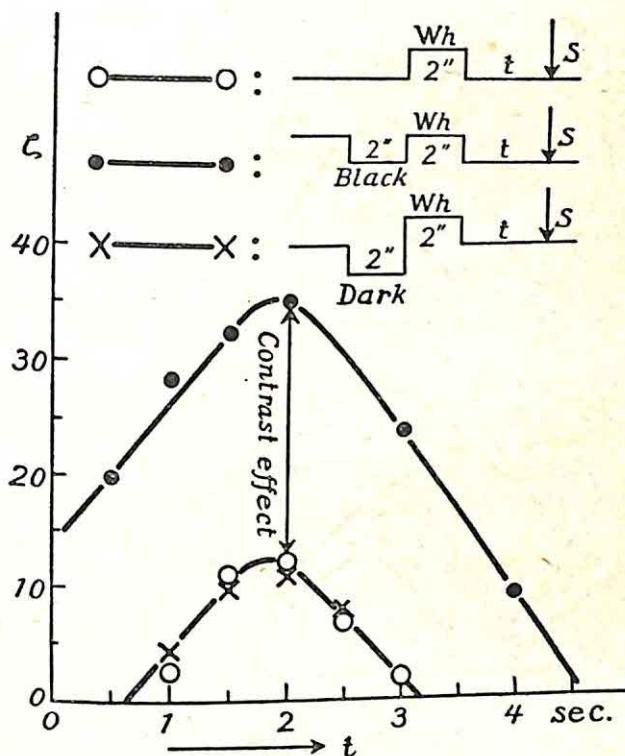
FIG. 18. EFFECT OF SUCCESSIVE BRIGHTNESS CONTRAST ON THE ζ-TIME CURVES. [MOTOKAWA (37).]

lus after a contrast situation had been set up. The insets in Fig. 18 and 19 show how the stimuli were ordered for brightness contrast.

Brightness Contrast

To study the effect of successive brightness contrast on electrical excitability, a 2° test spot was seen against the background of a screen held at .05 lux. The test spot was either black (no light) or white (84 lux). These were presented succes-

the crosses of the lower curve, had no effect. The amount of contrast was further shown to depend, somewhat, on the duration of the black stimulus, and more particularly, on the intensity of the adapting field. The contrast effect increased as the intensity of the adapting screen was raised, since raising the field level increased the "blackness" of the dark spot.

Simultaneous contrast was also investigated by placing a bright patch

within a dark, concentric ring. The effect of this was to raise ζ from 13 to 34.5. The difference in excitability in this instance was considered due to contrast.

Color Contrast

Successive color contrast is shown in Fig. 19, where it is seen that if

field of induction that could be mapped out by his electrostimulus method. Now he proceeded to demonstrate this in two papers on retinal induction (45, 48).

Optical Illusions

In the first report Motokawa measured the induced field surrounding
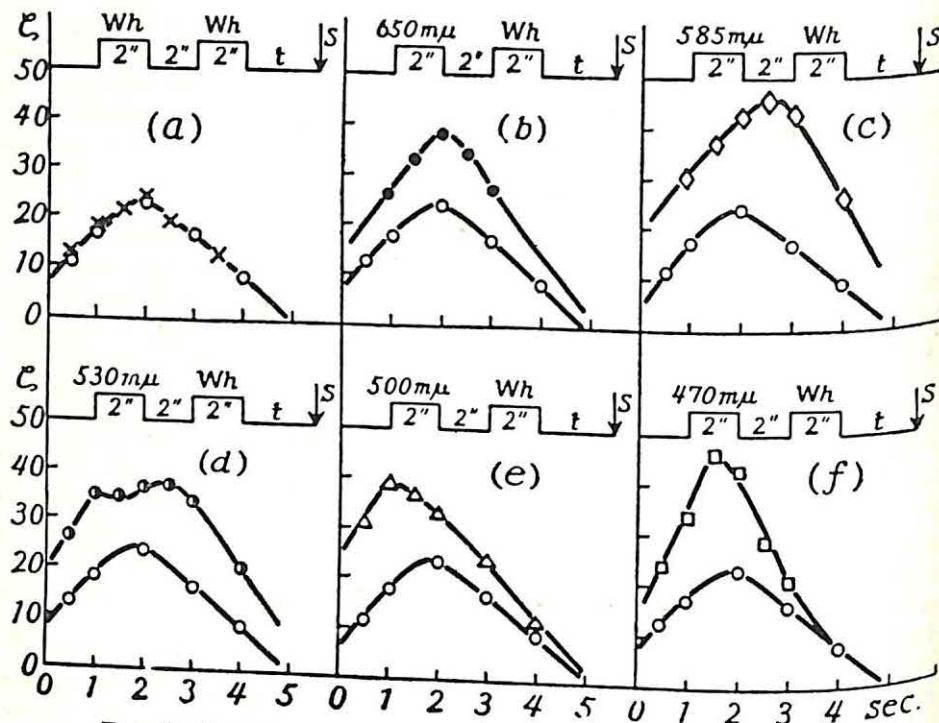
FIG. 19. EFFECT OF SUCCESSIVE COLOR CONTRAST. [MOTOKAWA (37).]

white light is preceded by colored light, the ζ maxima occur at times corresponding to the complement of the colored light used. White light, however, has no effect on the excitability curve of white light. This is seen in Fig. 19 (a). Further experiments showed the contrast effect to be dependent on the intensity of the inducing light, but not on that of the white test light.

Motokawa's supposition, as set forth in this paper (37), was that an optical image formed on the retina produced in the surrounding area a

geometrical figures (squares, triangles, and circles) and certain specialized configurations such as the Hering squares, the Wertheimer-Benary contrast figures, the Müller-Lyer illusion, and the Landolt ring (45). The procedure was to present an intense, yellow (sodium lamp) stimulus to the retina in the shape of the inducing figure. The exposure of this was for 2 sec. Then, after a delay of 3 sec., a small white test light was presented in different parts of the retina surrounding the inducing yellow exposure. This second exposure

was also for 2 sec. Finally, a 100-msec. electric pulse was applied to the eye after a second delay of 1.5 sec. (the ζ maximum for the Y process) for the purpose of finding the threshold phosphene. Contrast was defined as before; i.e., by the difference between the ζ-value determined for white light and the ζ-value determined after the above procedure.

The startling finding was that a gradient of field, as defined by this electrical measure of a contrast effect, existed around a given stimulus figure. This excitability field was strongest at the margins of the figure and fell off to zero with distance from the boundary. The rate at which the field strength declined depended upon where, with respect to the boundaries of the figure, the field was measured. Measures of field strength were used by Motokawa for constructing geometric representations of the gradients surrounding different figures. For example, the shape of the field surrounding a circle was circular, but the field surrounding a square was in the form of a cross, with the field falling off most sharply at the corners.

The same general procedure was then used to explore the retinal surrounds of the illusions mentioned above and it was shown that the fields of retinal induction were distinctively patterned by the figures involved. The suggestion was made by Motokawa that the experience of illusion may perhaps be accounted for by deformations in the fields of retinal induction surrounding a figure. These fields may be due either to the figure itself or to some other configurations in the vicinity. In either case, physiological processes serving the retinal image are warped or modified in a manner characteristic of the illusion.

Propagation of Induction Effects

This very fascinating work was continued in the second paper (48) on color contrast effects. Motokawa at this time introduced some new terminology. By *direct induction* he means the residual effect left in the *same* area of the retina by stimulation with colored light. It may be measured by his electrical method in the manner already described; i.e., by comparing ζ for white light alone with ζ for white light that has been preceded by colored light. So if yellow light is used, excitability will increase in the region of the ζ-time curve where the B process is, or, in this case, at 3 sec. Now if the yellow light is followed by the proper amount of exposure to blue light before the test with white light is made, the enhancement at 3 sec. disappears and the ζ-time curve is identical to the simple white light curve. This is called *neutralization*. If the intensity of the neutralizing blue light is too high, overneutralization occurs.

Motokawa then studied the effect of the intensity of the neutralizing light in direct induction, and the temporal position of the neutralizing light with respect to the inducing light. With regard to the latter, he found that the contrast effect of an inducing light persisted for 15 sec. or more and that the effectiveness of a neutralizing light increased the later it followed the inducing light.

By *indirect induction* is meant the residual process that exists in an area adjacent to the area stimulated. If the ζ-time curve is obtained using a white test-patch in an area next to where yellow light has acted, the maximum will be at 1.5 sec., not at 3 sec., as in the direct induction case where the same retinal area is involved. This effect may be neu-

J. W. GEBHARD
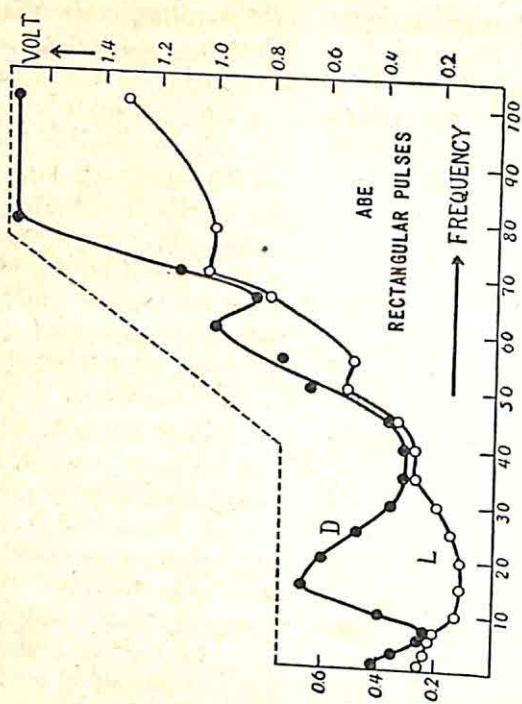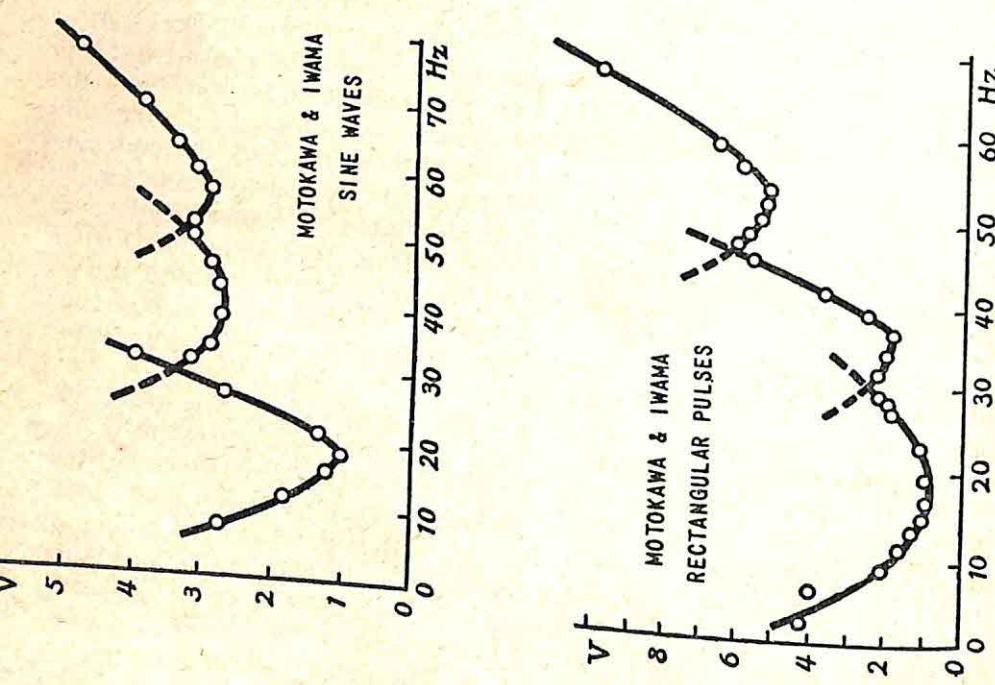
FIG. 20. STRENGTH-FREQUENCY RELATIONSHIP FOUND FOR THE LIGHT-ADAPTED EYE BY MOTOKAWA AND IWAMA (60) AND ABE (1). CURVES L AND D DENOTE LIGHT- AND DARK-ADAPTED EYES, RESPECTIVELY.

tralized by a light of the same color as the inducing one.

The previous work (45) had shown that retinal field gradients were rather extensive. Indeed, they were astonishingly large, sometimes reaching out from the area stimulated as far as 4° to 6°. By setting up an "inducer-detector" arrangement in terms of stimulus lights, Motokawa next showed that the effects of indirect induction were propagated over distances of 46° of arc in the retina. This propagation effect would not pass through the blind spot nor across a barrier formed by the image of a white light stimulus.

The whole area of vision that Motokawa has touched upon in the studies reported in this section has been one in which little real progress has been made, in spite of great interest, especially on the part of psychologists. The discovery that measurable alterations in the electro-physiology of the retina can be demonstrated to occur during contrast and illusion processes is most welcome. The verification of this should afford an engaging challenge to many workers in the field of vision.

IV. Stimulus Strength-Frequency Relationship

It has long been known from work on tissues other than the retina that optimum frequencies of excitation exist at which thresholds to intermittent electric stimulation are lowest (27). U-shaped curves result when threshold current (I) is plotted against log frequency (n) (7). The I-log n curves are symmetrical about the optimum frequency and their theoretical importance has been discussed in detail by Hill (21), Coppée (7), Katz (24), Schaefer (70), and others. Rohracher (69) showed in 1935 that the stimulus strength-frequency curve for just noticeable

flicker phosphenes passed through a minimum at about 20 cy./sec. These data were obtained with sine waves for the light-adapted eye. Schwarz (71, 72, 73, 74, 75) verified and extended Rohracher's observation that only one frequency of maximum stimulus effectiveness existed for the light-adapted condition. Motokawa and Iwama (60) and Abe (1), however, have reported three and five minima, respectively, for electrical thresholds during light adaptation, although in another place Motokawa reported data on two minima, while mentioning three and four (38). Abe, further, stated that the minima obtained during adaptation to darkness were not the same as those found during the light condition. Some of these results are shown in Fig. 20. It is interesting to note that, for Abe, the 20 cy./sec. minimum disappears altogether during dark adaptation. Motokawa considered this essentially a resonance phenomenon with the most conspicuous natural frequency of the eye to be at about 18 cy./sec. for the light-adapted condition. He offered, in support of this opinion, evidence that the period of oscillation into which the eye was thrown by a sensitizing electric shock was 54.85 msec. (44). This works out to be about 18.2 cy./sec. and corresponds well to the strength-frequency data.

This is the first, and to date the only, opportunity available to compare any of the findings of the Tohoku laboratory with those of other investigators who have recently looked into visual electrostimulation problems. It has already been mentioned that Rohracher and Schwarz found only one minimum in the strength-frequency curve. Gebhard (11, 12), using several different wave forms, also found but one unequivocal minimum. This was at about 20 cy./sec. as shown in Fig. 21.

The strength-frequency relationship for the dark-adapted eye was studied in considerable detail by Schwarz (75). Abe's findings for this condition are completely at variance with Schwarz' data. Nevertheless, nothing can be concluded from all this. Either there is only one minimum in the stimulus strength-frequency curve or there are more. There are objections to be raised to all of the investigations and it is evi-
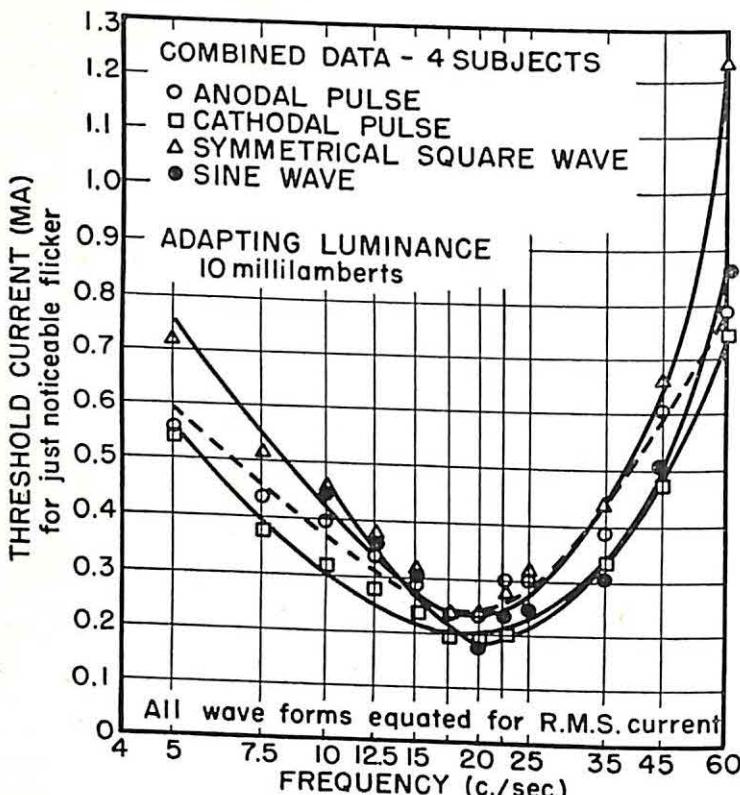
Fig. 21. Strength-Frequency Relationship for the Light-Adapted Eye.
[Found by Gebhard (12).]

Abe's supposition that the absence of the 20 cy./sec. minimum under dark adaptation was due to the low electrical excitability of the rods was pursued further by Tukahara and Abe (80). These authors obtained ζ-time curves on eyes pre-illuminated with intermittent white and colored lights of different intensities and at different positions on the retina. They concluded from their data that the natural resonant frequency of the rods, when light was acting on the eye, was 20 cy./sec. Unfortunately, dent that this problem requires restudy.

V. *Measurement of General Fatigue*

People have been trying for years to tease an indicator of fatigue out of photic flicker effects (6, 29, 76). The success of this effort has been small. Notwithstanding the poor record of using light, Motokawa and Suzuki (64) introduced in 1948 a method of measuring fatigue that made use of the electrically aroused

phosphenes. The procedure was to impress 20 cy./sec. pulses across the head in the manner by now familiar. Under a normal, rested condition, the thresholds S_1 and S_2 were obtained for the appearance and disappearance of flicker, respectively, by increasing and decreasing the stimulus voltage at a uniform rate. S_1 was found always to be higher, so the difference ΔS_o was gotten by subtracting S_2 from S_1. Therefore,

$$\Delta S_0 = S_1 - S_2.$$

This process was then repeated after fatigue to obtain the difference, ΔS. The comparison between ΔS_o and ΔS was made by finding their difference. $\Delta S - \Delta S_o$, therefore, is the measure of fatigue that is reported to be sensitive enough to be useful.

Correlation with Oxygen Deficiency

In a paper that appears to have been written earlier than the above, Motokawa and Iwama (58) reported work on electrical excitability and oxygen deficiency. Here also were used the appearance and disappearance thresholds for flicker and the difference between them. Measurements were taken under moderate light adaptation in a low pressure chamber over a range of pressures equivalent to over 20,000 feet of altitude. The main result was that while the means of the appearance and disappearance thresholds stayed about the same, the former rose somewhat while under low oxygen pressure, whereas the latter fell. The difference between them, then, turned out to be an indication of oxygen deficiency and can be seen to be like the ΔS term referred to above. The authors also measured the sensitivity of the eye to light during decreased oxygen pressure and concluded that

their new-found measure of electrical excitability was the more suitable.

Effect of Work

Suzuki (78) next carried the method over into the field of exercise and compared oxygen consumption with thresholds of electrical flicker. The electrical data before and after work were expressed in terms of $\Delta S - \Delta S_o$. The oxygen consumption was measured by conventional methods and was expressed as the difference

$$O - O_o$$

where O_o is the quantity of oxygen consumed per minute at rest and O is that consumed during work or in the course of recovery from work. In a series of experiments on the bicycle ergometer, running and stair climbing, the data from the two methods ran parallel to each other with remarkably small discrepancies. Suzuki concluded that the two methods were measuring the same thing and that $\Delta S - \Delta S_o$ was perhaps the more sensitive measure.

A point of difference that should be mentioned is that whereas Motokawa and Iwama (58) found the mean values of appearance and disappearance to stay the same for reduced oxygen pressure, Suzuki's results showed that S_1 and S_2 tended to change in the same direction, while at the same time their differences were changing. The *difference*, nevertheless, continued to be the critical feature of the measure.

Suzuki (77) also studied the course of $\Delta S - \Delta S_o$ during the daily activities of workers. Three classes of workers, chosen on the basis of the estimated amount of fatigue involved in performing their jobs, gave the results shown in Table 4. The data are not extensive, but there is little overlap. These findings offered

further evidence that Motokawa's method was measuring something useful and real about fatigue.

Research on the Method

It had been mentioned by Motokawa and Iwama (58) that the rate at which the appearance and disappearance thresholds were approached seemed to be important in establishing the value of the thresholds. The matter of time was not well controlled

TABLE 4

MAXIMUM DEGREE OF FATIGUE IN INDUSTRIAL WORKERS AS MEASURED BY MOTOKAWA'S METHOD
[AFTER SUZUKI (77)]

Severity of Work	Job Examples	Maximum Fatigue (Range of $\Delta S - \Delta S_o$ in mV)
Light ($N=4$)	Doctor, nurse, clerk, designer	196–216
Medium ($N=11$)	Locomotive engineer, telephone girl, telegraph operator, etc.	212–380
Heavy ($N=5$)	Stoker, coolie	432–635

in their experiment. The rate was controlled manually and the experimenter stopped varying the current at a signal from the subject. The experimenter's reaction time, as well as the subject's, undoubtedly caused a certain amount of overshooting. Mita, Abe, and Byonshik (32) carefully controlled such factors in a study in which they investigated the dependence of ΔS on the intensity of ambient illumination, the frequency of the flicker employed, and the rate of altering the stimulus current. They used a motor drive to control the rising and falling stimuli and a device that, at the subject's

signal, recorded where the stimulus value was at threshold. This eliminated the experimenter's errors, but, of course, not the subject's reaction time.

The results indicated that the measure ΔS could be analyzed into a factor, S, which is the mean threshold $(S_1 + S_2/2)$ and the reaction time, T, for the minimum stimulus. They proposed then, the equation

$$\Delta S = KS - 2TV,$$

where K is a constant and V is the rate of altering the stimulus current.

There are several psychophysical aspects of the data presented that require more study. For example, the frequency of the current pulses must be kept constant, since thresholds measured at values above and below 20 cy./sec. were observed to give somewhat larger measures of ΔS. This would appear to mean that the presence or absence of flicker can be more sharply discriminated at 20 cy./sec. than at any other frequency. Is there a significance in this for the strength-frequency relation? Another question is: why should threshold be proportional to the rate at which the current is raised or lowered only when adaptation is held constant? These are but a few of the interesting problems raised by this analysis. The only safe conclusion for the use of Motokawa's test of fatigue is that ΔS should be measured under carefully controlled conditions.

CONCLUSION

There can be little doubt that the reported work of Motokawa and his associates is exciting stuff. It represents a fresh and unique approach to vision that appears to promise great things. Unfortunately, this cannot be said without reservation.

The work of the Tohoku laboratory has come upon the rest of the scientific world so recently, and in such volume, that there has been little opportunity for visual workers to digest it, to say nothing of confirming it. An evaluation of Motokawa's findings cannot be made in any serious sense until a considerable amount of his experimentation has been redone, verified, or modified, as the case may be. For the present, one can only say that the work looks good. Indeed, it almost looks too good, in that the data of electro-analysis and that of more conventional experimentation fit together, in many respects, so beautifully.

It is possible, of course, to point out some obvious shortcomings. These were touched upon briefly in the introduction. The first has to do with the method of measuring the electrical thresholds. The reading of voltage is not the method of choice in this kind of work, although if proper precautions are taken, voltage measurements may be reasonably satisfactory. For example, a frequent redetermination of E_o is certainly needed to eliminate the consequence of voltage changes due to variations in resistance. Motokawa probably does this, although he does not make it clear that he does.

The second difficulty concerns the amount of data Motokawa has used in presenting his material. Complete tables of the measures obtained are never given. Consequently, no independent estimate of such matters as the variability of measures and the significance of differences is possible. No statements are made about how many measures were used to determine any given point on the numerous curves published. It can only be observed that the points plotted from the data generally can be fitted very well by smooth curves.

Data are always presented as typical examples of the measures obtained. The extent to which atypical data were collected cannot be judged. Finally, the number of subjects measured is rarely given. Since a more complete reporting of the data would do much to satisfy critics in many laboratories, this matter was referred to Professor Motokawa. In two personal communications[7] to the writer he makes the following observations. First, he stresses the importance of thoroughly trained observers. In most of his experiment he used two or three subjects, and two of the most frequently used have had seven years of experience. He notes that other workers in Japan have failed to reproduce his results until after adequate, long-term training had been given. In this connection, he attaches great weight to the ability of S to reproduce his thresholds. Since each threshold determination takes a rather long time, numerous measures cannot be obtained. A search for quick methods of determining the electrical thresholds yielded only unsatisfactory results. Finally, since he believes that there are many kinds of receptors in the retina, each with its own electrical threshold, great care must be taken to be sure the threshold obtained is the one associated with the receptor process under scrutiny. Hasty threshold determinations, some tapping one process and some another, yield values that belong to different populations and, hence, cannot be averaged in any meaningful way.

The total work, nevertheless, remains impressive, and these criticisms are in no way intended to detract from it. This review cannot be a substitute for reading the papers. It is hoped that it will merely serve

[7] February 5, 1952 and June 1, 1952.

two purposes: first, to arouse some interest in the work, and second, to urge those workers who have available facilities to verify the published results. Such study should not be difficult, since the technique is clearly reported and appears not to be particularly involved. If Motokawa's work stands, it will unquestionably be one of the most important contributions to the physiology of vision in many years.

BIBLIOGRAPHY

1. ABE, Z. Influence of adaptation on the strength-frequency curve of human eyes, as determined with electrically produced flickering phosphenes. *Tohoku J. exp. Med.*, 1951, **54**, 37–44.
2. ACHELIS, J. D., & MERKULOW, J. Die elektrische Erregbarkeit des menschlichen Auges wärhend der Dunkeladaptation. *Z. Sinnesphysiol.*, 1929, **60**, 95–125.
3. BOGOSLOVSKY, A. I. Über die Abhängigkeit der elektrischen Empfindlichkeit des Auges von den verschiedenen Adaptationsbedingungen. *v. Graefes Arch. Ophthal.*, 1935, **133**, 105–114.
4. BOUMAN, H. D. Electrical excitability of the eye. A study in clinical physiology. *Ophthalmologica*, 1940, **99**, 394–401.
5. BOURGUIGNON, G., & DÉJEAN, R. Double chronaxie du système optique de l'homme. *C. R. Acad. Sci., Paris*, 1925, **180**, 169–172.
6. BROZEK, J., & KEYS, A. Flicker fusion frequency as a test of fatigue. *J. industr. Hyg.*, 1944, **26**, 169–174.
7. COPPÉE, G. Stimulation by alternating current. *Cold Spr. Harb. Sympos. quant. Biol.*, 1936, **4**, 150–162.
8. CORDS, R. Über die Verschmelzungsfrequenz bei periodischer Netzhautreizung durch Licht oder elektrische Ströme. *v. Graefes Arch. Ophthal.*, 1908, **67**, 149.
9. EBE, M., ISOBE, K., & MOTOKAWA, K. Physiological mechanisms of color blindness. *Science*, 1951, **113**, 353–354.
10. FRÖHLICH, F. W. *Die Empfindungszeit.* Jena: G. Fischer, 1929.
11. GEBHARD, J. W. Electrical stimulation of the eye by sine waves, square waves and rectangular pulses. *Amer. Psychologist*, 1949, **4**, 358. (Abstract)
12. GEBHARD, J. W. Thresholds of the human eye for electric stimulation by different wave forms. *J. exp. Psychol.*, 1952, **44**, 132–140.
13. GERNANDT, B. Colour sensitivity contrast and polarity of the retinal elements. *J. Neurophysiol.*, 1947, **10**, 303–308.
14. GRANIT, R. Comparative studies on the peripheral and central retina. I. On interaction between distant areas in the human eye. *Amer. J. Physiol.*, 1930, **94**, 41–50.
15. GRANIT, R. The distribution of excitation and inhibition in single-fibre responses from a polarized retina. *J. Physiol.*, 1946, **105**, 45–53.
16. GRANIT, R. *Sensory mechanisms of the retina.* London: Oxford Univer. Press, 1947.
17. GRANIT, R. Neural organization of the retinal elements, as revealed by polarization. *J. Neurophysiol.*, 1948, **11**, 239–251.
18. HARTLINE, H. K. Nerve messages in the fibers of the visual pathway. *J. opt. Soc. Amer.*, 1940, **30**, 239–247.
19. HARTRIDGE, H. The visual perception of fine detail. *Philos. Trans.*, 1947, **232**, 519–671.
20. HECHT, S. Rods, cones, and the chemical basis of vision. *Physiol. Rev.*, 1937, **17**, 239–290.
21. HILL, A. V. Excitation and accommodation in nerve. *Proc. roy. Soc.*, 1936, **B119**, 305–355.
22. HIRONAKA, K. "Empfindungszeit" of an electrical phosphene and that of a light stimulus. *Tohoku J. exp. Med.*, 1950, **53**, 1–9.
23. IWAMA, K. Die elektrische Erregbarkeit des menschlichen Auges. *Tohoku J. exp Med.*, 1949, **50**, 71–77.
24. KATZ, B. *Electric excitation of nerve.* London: Oxford Univer. Press, 1939.
25. KLEITMAN, N., & PIÉRON, H. Loi de variation de la durée de la première phase dans l'établissement de la sensation pour des excitations lumineuses croissantes des cônes et des batonnets. *C. R. Soc. Biol. Paris*, 1924, **51**, 456–459.
26. KRAVKOV, S. V., & GALOCHKINA, L. P. Effect of a constant current on vision. *J. opt. Soc. Amer.*, 1947, **37**, 181–186.
27. v. KRIES, J. Ueber die Erregung des motorischen Nerven durch Wechselströme. *Ber. Verh. naturf. Ges., Freiburg i.B.*, 1884, **8**, 170–205.
28. v. KRIES, J. Die Gesichtsempfindungen.

Nagels Handb. Physiol., 1905, **3**, 109–282.

29. LEE, R. H., & HAMMOND, E. C. The effect of driving fatigue on the critical fusion frequency of the eye. *J. industr. Med.*, 1942, **11**, 360–363.

30. LE ROY. (Mémoire) Où l'on rend compte de quelques tentatives que l'on a faites pour guérir plusieurs maladies par l'électricité. *Mém. mat. phys. Acad. roy. Sci.*, Paris, 1755, 60–98.

31. LOHMANN, H. Ueber die Sichtbarkeitsgrenze und die optische Unterscheidbarkeit sinusförmiger Wechselströme. *Z. Sinnesphysiol.*, 1940, **69**, 27–40.

32. MITA, T., ABE, Z., & BYONSHIK, T. On the essential factors of Motokawa's method for measuring fatigue. *Tohoku J. exp. Med.*, 1951, **54**, 45–52.

33. MITA, T., HIRONAKA, K., & KÔIKE, I. The change in electrical excitability of the human retina caused by a flash of light. *Tohoku J. exp. Med.*, 1949, **51**, 379–388.

34. MITA, T., HIRONAKA, K., & KÔIKE, I. The influence of retinal adaptation and location on the "Empfindungszeit." *Tohoku J. exp. Med.*, 1950, **52**, 397–405.

35. MOTOKAWA, K. Retinal processes and their role in color vision. *J. Neurophysiol.*, 1949, **12**, 291–303.

36. MOTOKAWA, K. Physiological studies on mechanisms of color reception in normal and color-blind subjects. *J. Neurophysiol.*, 1949, **12**, 465–474.

37. MOTOKAWA, K. Physiological induction in human retina as basis of color and brightness contrast. *J. Neurophysiol.*, 1949, **12**, 475–488.

38. MOTOKAWA, K. On the mechanism of periodic excitability of nervous tissue. *Tohoku J. exp. Med.*, 1949, **50**, 307–318.

39. MOTOKAWA, K. Visual function and the electrical excitability of the retina. *Tohoku J. exp. Med.*, 1949, **51**, 145–153.

40. MOTOKAWA, K. Electrophysiological studies of color vision. *Tohoku J. exp. Med.*, 1949, **51**, 165–173.

41. MOTOKAWA, K. Spatial summation of optic stimuli in the human retina as revealed by electrical stimulation. *Tohoku J. exp. Med.*, 1949, **51**, 179–187.

42. MOTOKAWA, K. A physiological basis of color discrimination. *Tohoku J. exp. Med.*, 1949, **51**, 197–205.

43. MOTOKAWA, K. Physiological evidence for the three-components theory of color vision. *Tohoku J. exp. Med.*, 1949, **51**, 207–214.

44. MOTOKAWA, K. Periodic excitability of the human retina. *Jap. J. Physiol.*, 1950, **1**, 16–21.

45. MOTOKAWA, K. Field of retinal induction and optical illusion. *J. Neurophysiol.*, 1950, **13**, 413–426.

46. MOTOKAWA, K. Summation of the color processes in the human retina. *Tohoku J. exp. Med.*, 1950, **52**, 207–212.

47. MOTOKAWA, K. Selective inhibitory action of colored light upon the retinal color processes and its usefulness for analysis of the mechanism of color reception. *Tohoku J. exp. Med.*, 1950, **52**, 213–221.

48. MOTOKAWA, K. Propagation of retinal induction. *J. Neurophysiol.*, 1951, **14**, 339–351.

49. MOTOKAWA, K., & EBE, M. Scotopic process and on-elements in human retina. *Tohoku J. exp. Med.*, 1951, **54**, 215–221.

50. MOTOKAWA, K., EBE, M., ARAKAWA, Y., & OIKAWA, T. Studies on the physiological color blindness of the human fovea with the polarization method. *Jap. J. Physiol.*, 1951, **2**, 50–59.

51. MOTOKAWA, K., EBE, M., ARAKAWA, Y., & OIKAWA, T. Studies of rod-process by polarization method. *J. opt. Soc. Amer.* 1951, **41**, 478–481.

52. MOTOKAWA, K., EBE, M., ARAKAWA, Y., & OIKAWA, T. Retinal colour responses to microstimulation. *Nature, Lond.*, 1951, **167**, 729–730.

53. MOTOKAWA, K., & IWAMA, K. (This paper was to have appeared in the Volume 11 of the now discontinued *Jap. J. med. Sci. Biophys.* The subject matter was probably that of Reference 44.)

54. MOTOKAWA, K., & IWAMA, K. Electric stimulation of the human retina with exponentially increasing currents. *Tohoku J. exp. Med.*, 1949, **50**, 25–37.

55. MOTOKAWA, K., & IWAMA, K. Color receptors in frog eyes as revealed by electrical stimulation. *Tohoku J. exp. Med.*, 1949, **50**, 240.

56. MOTOKAWA, K., and IWAMA, K. The three color process in the retina of frog. *Tohoku J. exp. Med.*, 1949, **50**, 258.

57. MOTOKAWA, K., & IWAMA, K. Experiments of color contrast on excised frog eyes. *Tohoku J. exp. Med.*, 1949, **50**, 292.

58. MOTOKAWA, K., & IWAMA, K. The electric excitability of the human eye as a sensitive indicator of oxygen deficiency. *Tohoku J. exp. Med.*, 1949, **50**, 319–328.

59. MOTOKAWA, K., & IWAMA, K. The rela-

tion between the intensity of light and the electrical excitability of the human retina. *Tohoku J. exp. Med.*, 1949, 51, 155–164.

60. MOTOKAWA, K., & IWAMA, K. Resonance in electrical stimulation of the eye. *Tohoku J. exp. Med.*, 1950, 53, 201–206.

61. MOTOKAWA, K., & IWAMA, K. Color processes and physiological induction in frog's retina. *Tohoku J. exp. Med.*, 1951, 53, 341–349.

62. MOTOKAWA, K., IWAMA, K., & ENDŌ, T. Über den Einfluss unterschwelliger elektrischer Reizung des Auges auf die Lichtempfindlichkeit der Netzhaut., *Tohoku J. exp. Med.*, 1948, 49, 331–338.

63. MOTOKAWA, K., IWAMA, K., & TUKAHARA, S. Color processes in single retinal elements. *Tohoku J. exp. Med.*, 1951, 53, 399–406.

64. MOTOKAWA, K., & SUZUKI, K. A new method for measuring fatigue. *Jap. Med. J.*, 1948, 1, 200–206.

65. MOTOKAWA, K., & SUZUKI, K. Electrophysiological studies of color-blindness. *Tohoku J. exp. Med.*, 1950, 52, 195–206.

66. MOTOKAWA, K., & SUZUKI, K. Quantitative aspects of retinal inhibition. *Tohoku J. exp. Med.*, 1950, 52, 341–348.

67. MOTOKAWA, K., & SUZUKI, K. Analysis of the retinal processes of selective inhibition. *Tohoku J. exp. Med.*, 1950, 52, 349–359.

68. NAGEL, W. A. Einige Beobachtungen über die Wirkung des Druckes und des galvanischen Strömes auf das dunkeladaptierte Auge. *Z. Psychol. Physiol. Sinnesorg.*, 1904, 34, 285–290.

69. ROHRACHER, H. Über subjektive Lichterscheinungen bei Reizung mit Wechselströmen. *Z. Sinnesphysiol.*, 1935, 66, 164–181.

70. SCHAEFER, H. *Elektrophysiologie.* Vienna: Franz Deuticke, 1940–1942, 2 vols.

71. SCHWARZ, F. Über die Wirkung von Wechselstrom auf das Sehorgan. *Z.*

Sinnesphysiol., 1936–1938, 67, 227–244.

72. SCHWARZ, F. Über die Reizung des Sehorgans durch niederfrequente elektrische Schwingungen. *Z. Sinnesphysiol.*, 1939–1940, 68, 92–118.

73. SCHWARZ, F. Quantitative Untersuchungen über die optische Wirkung sinusförmiger Wechselströme. *Z. Sinnesphysiol.*, 1940, 69, 1–26.

74. SCHWARZ, F. Über die Reizung der Sehorgans durch doppelphasige und gleichgerichtete elektrische Schwingungen. *Z. Sinnesphysiol.*, 1941, 69, 158–172.

75. SCHWARZ, F. Über die elektrische Reizbarkeit des Auges bei Hell-und Dunkeladaptation. *Pflüg. Arch. ges. Physiol.*, 1947–1948, 249, 76–86.

76. SIMONSON, E., & ENZER, N. Measurement of fusion frequency of flicker as a test of fatigue of the central nervous system; observations on laboratory technicians and office workers. *J. industr. Hyg.*, 1941, 23, 83–89.

77. SUZUKI, K. Professional differences of fatigue as revealed by the method of electric flicker. *Tohoku J. exp. Med.*, 1950, 52, 1–7.

78. SUZUKI, K. Oxygen consumption during and after exercise and its relation to the degree of fatigue as measured by the method of electric flicker. *Tohoku J. exp. Med.*, 1950, 52, 9–16.

79. TUKAHARA, S. Modulators and dominators of the toad's retina. *Tohoku J. exp. Med.*, 1951, 54, 11–20.

80. TUKAHARA, S., & ABE, Z. Resonance phenomena of photopic and scotopic receptors. *Tohoku J. exp. Med.*, 1951, 54, 189–196.

81. VERRIJP, C. D. L'influence de l'adaptation a l'obscurité sur l'excitabilité de l'oeil humain. *C. R. Soc. Biol. Paris*, 1925, 93, 55–58.

SUPPLEMENT

Four new papers from Motokawa's laboratory have appeared since this review was written. Two of these report experiments on human subjects and two concern the application of the electrostimulation method to the visual processes of cats.

Color diagrams. The first paper on humans by Arakawa and Oikawa (1) describes the construction of a color diagram from the three physiological sensation curves. It will be remembered that these curves are found by measuring ζ for each wave length when the time after pre-illumination is held at 1, 2, and 3 sec. The authors point out that the color diagram calculated from the ζ-λ data bears a close resemblance to diagrams constructed from color mixing results.

Strength-frequency problem. In the second paper Motokawa and Ebe report a new procedure that grew out of the work on the stimulus strength-frequency relationship (5). They found that the frequencies of alternating current that produced minima in the strength-frequency curves also produced the same effects as colored lights in the retinal induction situation. In retinal induction, pre-illumination by blue light, for example, followed by white light shifted the maximum of the ζ-time curve to the locus normally occupied by the yellow. Now it was shown that prestimulation by alternating current of 36 cps followed by white light produced the same shift to the yellow locus as the blue light. This relationship was worked out to indicate a functional correspondence between 33–37 cps and blue, 40–45 cps and green, 47–55 cps and yellow, and 60–100 cps and red. The conclusion reached was that the stimulus strength-frequency curve is a resonance curve revealing the composite nature of the retinal receptors.

A more limited conclusion about these matters was reached by Meyer-Schwickerath (3, 4) who found evidence that central and peripheral processes could be distinguished by electrostimulation. There is a minimum in each of Meyer-Schwickerath's curves. One is at 20 cps for the peripheral curve, and the other is at 33 cps for the central curve. Motokawa mentions this as supporting his and Ebe's findings. In the main, however, Meyer-Schwickerath's curves are unlike those found by other investigators. A recent paper by Bouman, ten Doesschate, and van der Velden (2), on the other hand, reports data in support of a single minimum in the strength-frequency curve.

Evidence from cats. The first paper on cats by Motokawa, Iwama, and Ebe (6) presents supporting evidence that

properly chosen frequencies of alternating current play the same role in induction as colored lights. The correspondence between frequency and hue turns out to be about the same for men and cats.

The last paper is by the same authors (7). Using the method found successful with frogs and toads, they report that the ζ-time curves for cats passed through maxima at 1, 1.5, 2.25, and 3 sec. when the pre-illumination was red, yellow, green, and blue light, respectively. These times became longer when the body temperature of the animals fell during the experiments, suggesting that temperature may relate the mammalian response times to the more sluggish reactions of the amphibians.

SUPPLEMENTARY BIBLIOGRAPHY

1. ARAKAWA, Y., & OIKAWA, T. Chromaticity diagrams and retinal color processes. *Tohoku J. exp. Med.*, 1952, 56, 291–297.
2. BOUMAN, M. A., TEN DOESSCHATE, J., & VAN DER VELDEN, H. A. Electrical stimulation of the human eye by means of periodical rectangular stimuli. *Documenta Ophthalmal.*, 1951, 5–6, 151–168.
3. MEYER-SCHWICKERATH, G. Unterschiedliche elektrische Erregbarkeit zentraler und peripherer Netzhautfasern. *Ber. dtsch. opthalmal. Ges., Munich*, 1950, 56, 70–73.
4. MEYER-SCHWICKERATH, G., & MAGUN, R. Über selektive elektrische Erregbarkeit verschiedener Netzhautanteile. *v. Graefes Arch. Ophthalmal.*, 1951, 151, 693–700.
5. MOTOKAWA, K., & EBE, M. Selective stimulation of color receptors with alternating currents. *Science*, 1952, 116, 92–94.
6. MOTOKAWA, K., IWAMA, K., & EBE, M. Color processes caused by alternating currents in the mammalian retina. *Tohoku J. exp. Med.*, 1952, 56, 215–222.
7. MOTOKAWA, K., IWAMA, K., & EBE, M. Retinal color processes in cats. *Jap. J. Physiol.*, 1952, 2, 198–207.

THE SZONDI TEST: A REVIEW AND CRITICAL EVALUATION

L. J. BORSTELMANN
University of North Carolina   AND

W. G. KLOPFER
*Norfolk, Nebraska State Hospital and
University of Nebraska Medical School*

The appearance of Deri's *Introduction to the Szondi Test* in 1949 (15) made available an organized body of theoretical and interpretive postulates for this technique. Szondi's *Experimentelle Triebsdiagnostik* appeared in 1947 (40) but an English translation has become available only recently (41). Thus Deri's text has been the primary reference for the test in the United States.

The Szondi Test, as any new proposition, in its exploratory phase has tended to produce isolated research with a minimum of communication among investigators. The *Szondi Newsletter* (45) established in 1949 cannot fulfill the coordinating function for which it was designed since most investigators seek other outlets to publicize their work. This state of affairs makes the present status of the technique somewhat ambiguous.

The problems of mediational rationale, validation, and personality theory inherent in the Szondi Test would seem to be in many ways representative of those generally encountered in multidimensional personality measures. A careful scrutiny of these problems should provide some general methodological principles for the investigation of projective techniques.

In the present evaluation of the Szondi Test, the rationale will first be examined in terms of the basic assumptions regarding the processes of test behavior and their psychological significance. These assumptions will then be evaluated in the light of the available research. Finally, there will be a general discussion of the theoretical and methodological problems characteristic of techniques such as the Szondi Test.

THE TEST

The administration and interpretation of the Szondi Test are presented extensively by Deri (15). A brief description will suffice for the present purpose.

Administration and Scoring

The Szondi Test materials (42) consist of 48 facial photographs of European psychiatric patients, with six representatives of each of eight diagnostic categories: passive male homosexuals, male sadistic murderers, epileptics in the intraparoxysmal phase, hysterics, catatonic schizophrenics, paranoid schizophrenics, manic-depressives in the depressive and in the manic phases. The test pictures are arranged into six sets, each containing one picture from each of the eight diagnostic groups.[1] The sets are administered successively to the subject with instructions to select the two "most liked" and the two "most disliked" pictures from each set. The subject thus has six opportunities to like, dislike, or ignore each class of pictures. The resulting pattern of 24 choices (12 likes and 12 dislikes) is represented by an eight-dimension profile. Each category is scored in terms of intensity (four to six choices is a "loaded" score and fewer choices yields an "unloaded" score) and direction of selection (a two to one imbalance of like and dis-

[1] Position of a given category within the sets is randomized.

like choices is a "plus" or "minus" score, four or more choices without such imbalance is an "ambivalent" score, and less than two choices in both directions is an "open" score).

Interpretive inferences from a single administration are considered tenuous and inadvisable. The standard procedure recommended is ten separate administrations with at least daily intervals.

Basic Assumptions

Szondi originally devised the test in connection with his theory of genetically determined personality (43). He postulated that the eight diagnostic syndromes of the test represent basic drives which are hereditarily determined. Both manifest physical appearance and personality are considered to be a function of genetic structure. The pictured patients have a definitely diagnosed psychopathology and, therefore, a known genetic structure and personality. The subject responds to the test pictures in terms of the relations of his own genetically determined basic drives to those of the pictured patients. The picture selections of the subject thus provide a basis for inference about characteristics of his personality. Szondi's constitutional orientation to personality probably accounts for the summary dismissal of the technique by many psychologists in this country. However, the merits of his genetic formulation need not be argued here since this rationale is not deemed essential to the technique.

Deri, a student of Szondi, seems to ignore his theory that personality characteristics are genetically determined, but continues to maintain that personality characteristics are reflected in the physiognomy. The diagnostic syndromes represented in the test are considered to be extreme manifestations of personality dimensions present within all persons. The subject's affective selections among the pictures are then due to interaction between eight need-systems within the individual and these needs as represented in the pictures. Thus, the homosexual pictures represent the "tender love" needs, the sadist pictures represent the "aggressive love" needs, etc. The basis of the subject's response seems to be essentially nonconscious in that he reacts in terms of stimulus qualities of which he is unaware.

As Schafer (35) has pointed out, Deri presents a body of interpretive hypotheses with a minimal elucidation of the mediational processes existent between stimulus and response. Deri makes the general assumption that the psychopathological dynamics of the various diagnostic syndromes are somehow communicated to the subjects through the medium of the test pictures. This general assumption embodies a number of important considerations.

The dynamics of diagnostic syndromes are assumed to be well established, with essential agreement as to clearly differentiating characteristics. These postulated dynamics are further assumed to be identical and equally well represented in all patients having a given diagnosis. Such faith in nosological clarity is not generally maintained in psychiatric practice and research.

That certain internal psychological states are reflected differentially in physiognomy has not been demonstrated. Deri seems to have created a paradox in adapting the test to a nongenetic formulation. In rejecting Szondi's genetic postulates she has dispensed with the assumption of constitutional origin, which is the essential link between personality and physiognomy. Although Deri's formulation of mental illness and personality has been more generally ac-

cepted than Szondi's, the role of physiognomy in her schema is more obscure.

The assumptions of nosological clarity with unique physiognomic manifestations are essential to Deri's hypotheses of the significance of reactions to various picture categories. The assumption is that each category has a unique stimulus value, an implicit and unverbalized associated meaning which forms the basis of the subject's affective reactions. These associative stimulus values must be generally constant in order to warrant the use of comparative inference.

Another basic assumption is that the subject makes his picture selections on the basis of personal affective reactions to the associated stimulus qualities. The forced-choice structure of the test defines individuality in terms of particular selection from limited alternatives. The existence of any cultural bias in favor of certain responses will tend to produce stereotyped affective reactions rather than choices of individual significance. Therefore, the pictures within any test set should have essentially equivalent group affective values.

Personality Inference

The essence of validity for any projective technique is the descriptive or predictive accuracy of the interpretive hypotheses derived from the test behavior. The validity of such inferences can only be demonstrated when the hypotheses are stated in a testable form. Are the Szondi hypotheses capable of systematic investigation?

Deri sets forth certain general hypotheses about test behavior. Frequency of selection within a given category of pictures is considered a measure of the extent of tension within the need-system represented. Thus, the absence of selections in a given category indicates an absence of need tension due to some immediate release, or a more enduring tendency to satisfy the need behaviorally. Conversely, a high frequency of selections within a category indicates an existing tension within the need-system. The role of this need-tension in the personality will be reflected in the manner of selection. Predominant "like" selection within a category is due to some process of identification with the need, or acceptance of it. Predominant "dislike" selection is then due to some kind of counteridentification or rejection of the particular need. The presence of both "like" and "dislike" choices within a category represents internal conflict in regard to the need.

The foregoing hypotheses apply to the interpretation of a single profile and form the basis for the more important interpretive emphasis upon the consistency or variability of choice patterns within a series of administrations. The Szondi Test is unique in this focus upon the personality as a process. Categories with unstable selection patterns represent the more unstable areas of personality in which symptom formation is to be expected. Conversely, consistency of test behavior is to be expected only in those categories which are representative of the stable, core or basic need-systems within the subject. The test is thus not amenable to the traditional measures of reliability, but requires rather the systematic study of unreliability.

The body of interpretive hypotheses for a given technique is an attempt to relate the particular test behavior to more generalized concepts of behavior or personality. The

schema set forth by Deri presents certain difficulties in this regard. The interpretive hypotheses are couched in the language of depth psychology without specification of the phenotypical referents of the concepts involved. Knowledge of their relationships to observable behavior appears essential to validation. Herein lies the clue to that slippery, evasive argument of pragmatic validation that the clinician sometimes employs, i.e., "it works." The vague, ambiguous, and generalized nature of the Szondi hypotheses results in all things being true of all people. The interpreter is often left with propositions having multiple application rather than idiosyncratic reference. The very plastic nature of case material will tend to support such generalized statements. Of course, there is the argument that individual interpretation is actually a process of ordering data in terms of known classes and that, therefore, the concept of a "unique" personality is really a myth. Nonetheless, a technique which purports to describe individuality must describe the individual in terms amenable to validation by external criteria. The need for an explicit statement of the relationships of hypothesized dynamics to overt behavior in given situations has been well stated by Ainsworth in her recent discussion of the validation problems of projective techniques (1).

REVIEW OF RESEARCH

The present review is intended as an evaluation of the Szondi Test in terms of the studies purportedly investigating either the basic rationale or the interpretative hypotheses of the technique. For a comprehensive bibliography, including expository descriptions and case illustrations, the reader is referred to David (12).

Validation of any instrument for personality assessment such as the Szondi may be approached in essentially two ways: (a) A trial-and-error method of empirical investigation can be employed, whereby the body of interpretive hypotheses is gradually verified, altered, or rejected.[2] Obviously a tortuous process, requiring patient exponents and critics, this method has the inherent pitfall of convenient memory. The "correct" hypotheses are readily reinforced, but the "incorrect" ones tend to die a lingering death without any adequate diagnosis of the ailment. This approach has characterized much of the research to date on projective techniques. (b) The problem can be attacked at its source, namely, the adequacy of the basic rationale underlying the body of interpretative hypotheses. Note that this approach does not eliminate the necessity for empirical validation, but is concerned with the prior question of whether the technique is actually capable of validation. The writers' point of view is that any technique purporting to reflect personality must have hypotheses based upon a demonstrable rationale. Accordingly, those studies dealing with the process of interaction between subjects and the Szondi materials will be considered prior to the studies of clinical validation. In this way any evident limitations of rationale may help to illuminate the results of inquiry into the interpretive hypotheses.

The only previous review of the Szondi literature is a brief survey by Guertin and McMahan (22). All research covered in that review is considered in the present appraisal.

[2] Sometimes referred to as the method of successive clinical predictions (33).

Studies of Basic Rationale

Investigations of the mediational assumptions inherent in the Szondi Test have dealt mainly with two central problems: the adequacy of the test pictures for communication of personality characteristics, and the assumption that picture selection is a function of individuality.

Content values of the test pictures. As previously noted, a number of subassumptions are involved in the central mediational assumption of the technique, which is that the subject reacts to the personality of psychiatric patients through the medium of facial photographs.

Fosberg (17) and Rabin (31, 32) have studied one aspect of this problem, the identification of diagnoses from the pictures. The results are in essential agreement and are best illustrated by Rabin. He found that both untrained and trained observers did significantly better than chance, with the latter group superior. Naive subjects tended to improve with increased knowledge of psychiatric disorders. The initial significance and subsequent improvement are both due largely to correct identification of the homosexual and manic pictures. Rabin concludes that physiognomic clues are important, that there is a need to calibrate the observer's ability, and that the findings tend to support the Szondi pictures as meaningful stimuli.

Identification by the use of diagnostic labels is appropriate as a study of the Szondi rationale of associated meanings only in so far as the dynamics implied by the labels are clearly understood by the observer as the hypothesized ones. The existing semantic confusion, ambiguity, and variability of nosological classifications do not appear to be carefully resolved in the present instances.

Rather, the Szondi pictures have here been examined for their appropriateness to the observer's unknown nosological frame of reference.

Szondi reports that the communicability of the particular test pictures was established by means of free associations to a wide sampling of pictures for each category. Those producing responses most appropriate to the category rationale were then selected for the test. Deri illustrates this procedure somewhat loosely in her chapter on "Experiment in Factorial Association" (15, pp. 17–24). Unfortunately, she gives no indication of any control for rater bias.

Wallen (44) has studied the stimulus values of the catatonic (k) and paranoid (p) pictures by having subjects select the most appropriate characteristics for each picture from a list of thirteen trait pairs. He found no difference between the two categories as to the proportion of votes given to any trait, but there was some suggestion that k's are seen as more "tense" and "inferior," and less "dominant" than p's. The pictures were reclassified for further analysis according to muscular tension. Those perceived as contracted were described as more "self-centered" and the relaxed ones as more "kind" and "optimistic." The latter group of pictures showed greater "like" selection in a related study. Wallen concludes that the Szondi Test requires a frame of reference based upon learned approach-avoidance tendencies in regard to given facial characteristics. This study cuts to the core of the problem of appropriate stimulus values, although it is somewhat limited by the various meanings which may be assigned to trait names by different judges. However, the results do question the adequacy

of the assumption that all pictures within a category have a given meaning.

Klopfer and Borstelmann (27) repeated Szondi's experiment of free association to the pictures and supplemented this procedure by matching the pictures to personality descriptions for each of the eight diagnostic categories. Interjudge agreement was 80 to 90 per cent for rating of the associations in terms of appropriateness to the eight dimensions of the test. Subsequent panel decisions resulted in associative valences[3] which agreed with those derived from matchings by subjects in 41 out of 48 pictures. However, only 22 of these valences were those hypothesized by Szondi. The empirically designated "homosexual" and "manic" pictures showed very high agreement with those designated by Szondi, while the epileptic pictures agreed very poorly (out of nine empirically derived valences, only one picture was actually an epileptic). The investigators concluded that the present stimulus materials of the test do not satisfy the basic assumption of appropriate communality of meaning within categories. A redesignation of the pictures in accordance with empirically demonstrated association values was suggested.

There has been much discussion and some study of the bases of picture selection. Some have suggested that the perceived nationality, dress, facial expression, etc. may influence picture choices. This is in contrast with the Szondi-Deri view that the category of mental disease is the determining factor in picture selection. David (11) questioned his subjects

(paranoid schizophrenics and student nurses) about the identity and nationality of the pictured patients. He found that the psychotics were more prone to accept the pictures as familiar personages than were the nurses. Both David's procedure and his discussion of results fail to deal definitively with his stated problem of peripheral versus dynamic bases of selection. The findings of this study do not indicate just what factors actually determined the selections of the subjects.

Blessing, et al. (3) have investigated the specific effects of clothing and background characteristics of the Szondi pictures upon affective selection. They found that elimination of all photographic elements except the head produced no change in selection patterns for group administration to college students, but resulted in significant group changes for individual administration to psychiatric patients. This study is an excellent example of the confused and nondefinitive experimental design that is frequent among studies of projective techniques. Different subjects are used under different conditions with no explanation or attempt to reconcile these uncontrolled elements. An explanation of the obtained results in terms of clothing differences among the categories completely misses the main problem—the meaning that these characteristics convey to the subjects. The assumption that the stimulus values hypothesized by Szondi are valid is gratuitous.

In summary, there is positive evidence that some of the test pictures do communicate the qualities hypothesized by Deri, especially in the homosexual and manic categories. The specific elements of mediation are unknown. Only two of the eight Szondi categories seem to have the stimulus values appropriate to the

[3] Associative valence is a term used to refer to the meaningful cluster of associative elements which are produced by a specified population in response to a given picture or group of pictures.

rationale. The failure of the remaining categories to communicate appropriately raises serious questions regarding the use of the test as presently constituted. The suggestion has been made that materials may be restructured in terms of empirically demonstrated stimulus values. Thus, the crucial point is whether subjects in general respond to a given picture as of manic quality, not whether the patient pictured is actually a manic or not. Finally, there is some evidence that such elements as facial muscular tension may contribute to particular associated meanings.

Picture selection and individuality. In addition to the assumption of appropriate content values there is the assumption that picture selections are due to individual reactions to the content values. Harrower's (23) collection of some 1300 profiles from neurotic and nonclinical subjects indicates the existence of popular selections, most of them common to both types of subject. Such stereotypes of selection must be explained either by positing a communality of need structure within the group or by attributing the choices to culturally biased reactions to certain physiognomic or pictorial characteristics. In order that the test provide a maximal opportunity for the expression of individuality within the group, the alternatives for selection should have essentially equivalent group affective values.

Since interpretation is based upon reactions to categories of pictures, all pictures of a given category are assumed to have equivalent selective significance. Guertin (19) found that selections from a single set were highly reliable, but reactions to the same category in different sets showed no better than chance agreement. He concludes that the subjects respond to specific pictures rather than to categories as presently hypothesized.

The inadequacy of the assumption of intracategory equivalence was further demonstrated by Guertin (21) in a factor analysis of preference ratings for two pictures from each category. The resulting factors showed no relationship to test categories, indicating that none of the category pairs showed consistent preferential reactions.

An interesting approach to the assumption of intracategory equivalence has been employed by Lubin and Malloy (28). They hypothesized that selections of pictures of the same category should show more positive relationships than pictures from different categories. Single protocols were obtained from 100 psychiatric patients. All possible pairs of pictures from each category were compared in a chi-square analysis of like, dislike, or not selected. No category had all members positively related, while all had certain positive pair relations, although no more than might be expected by chance. The conclusion was that the assumption of intracategory equivalence is not tenable.

The problem of affective variability of the stimuli has also turned up in studies by Guertin (20) and Davidson, *et al.* (14) comparing selections of Szondi pictures with culturally more familiar pictures. Guertin found that roughly comparable United States pictures evidenced variability equivalent to the Szondi pictures when independently ranked as to preference. He concludes that some of the stimuli are diagnostically better than others since they produce greater variability. Davidson, *et al.* report similar results, demonstrating certain popular reactions for both Szondi and control pictures.

Variability of preference ranking would seem to be a desirable but not conclusive criterion for stimulus structure. Equivalent central tendencies are also essential. The essential structural requirement is equipotentiality *within each set*, not equivalence of all test pictures since each test set is given separately. Therefore, only intraset comparison will answer this question. The contrast with non-Szondi pictures seems irrelevant unless associative stimulus values are considered. Guertin's "control" pictures actually show much less variability when compared directly with the Szondi pictures. However, Guertin's rankings of the Szondi pictures do provide a basis for restructuring of stimulus materials since the present arrangement proves inadequate.

The two most definitive studies of affective stimulus values are reported by Szollosi, *et al.* (39) and Borstelmann and Klopfer (5). The Szollosi group analyzed single profiles from a heterogeneous sampling of the normal, i.e., nonclinical, population for individual picture reaction patterns. They found that the six pictures of any given test category were not similar enough in affective selection to justify the assumption of equivalence among category members. The limitations of a forced choice situation where the stimuli are of unequal strength are pointed out in that interpretive inferences are demanded even though the individual reactions are equivalent to group stereotypes. They conclude that the stimulus materials should either be restructured for category interpretations or else used for reactions to individual pictures.

Borstelmann and Klopfer demonstrated that individual picture reaction values are not a function of some inherent stimulus quality (as assumed by Deri) but are actually relative values dependent upon the stimulus configuration of a particular test set. Affective ratings of the 48 individual test pictures by 400 college students revealed no consistency in median affective values within test sets or among members of any category. The test sets were then reconstituted so as to more nearly approach equivalence of affective stimulus values. Thus, the most liked pictures of each category were assigned to one set, the most disliked to another set, etc. The popular trends of selection were significantly reduced with the reorganized test materials. However, categories still differed markedly from each other in the general level of affective pull.

This raises a fundamental, and possibly insoluble, problem in the use of the forced-choice technique for the Szondi pictures. Szollosi, *et al.* infer that discrepancy in preference values indicates differential picture qualities. Wallen found that distinguishable stimulus qualities were related to divergent affective values. Therefore, existing evidence suggests that the very stimulus qualities that produce distinguishable associations also produce divergent affective reactions. If so, then the forced-choice structure will not reflect the desired individuality of response, since the alternatives for selection are unequally biased. Of course, both the Wallen and the Borstelmann and Klopfer findings are limited to the circumscribed population of college students. But Harrower's studies with noncollege groups show similar selection stereotypes. Even if the results should subsequently prove to be due to the peculiar need structure of college students as a group, the detection of individuality within the group is still questionable. Thus the present test structure would

yield only group, not individual, identification.

In summary, the necessary structural requirements of the Szondi forced-choice technique as a device for the reflection of individual uniqueness do not seem to be adequately fulfilled at present, nor even capable of fulfillment. Such a framework assumes that selection of pictures is due to the particular need structure of the individual. To the extent that responses are determined by group bias, the expression of individuality is necessarily restricted. The presence of persistent "picture pulls" within the present test structure has been clearly demonstrated. Furthermore, the establishment of equivalent selective significance among the test categories seems highly improbable since associative and affective qualities are not independent. Thus, pictures of demonstrated "manic" quality are generally liked while those of "sadistic" quality have generally negative values.

Studies of Interpretive Hypotheses

The studies reviewed above have not been concerned with interpretive hypotheses, but only with the question of whether the technique is capable of validation. With the limitations and deficiencies of the Szondi rationale now set forth, the studies dealing with interpretive hypotheses may be more rigorously appraised.

Normative studies. Deri makes many statements of genetic and socioeconomic trends as illustrations of interpretive hypotheses. A number of studies have reported picture selection patterns for various groups. Hill (25), Guertin (18), and Fosberg (17) have analyzed Szondi Test behavior in terms of deviations from chance phenomena. Cohen (7) provides an excellent didactic presentation of the chance probabilities for

each type of category selection (plus, minus, etc.) with any given frequency of choices for the category.

Guertin compared profiles of hospital attendants with those derived by random numbers. He found no difference in frequency of category choice or in positive selection, more negative selection and intracategory imbalance for the subjects, and more ambivalent selection within categories for the chance profiles. He concludes that frequency of selection is not indicative of need-tension, that ambivalence is indicative of stability rather than conflict, and that negative selection is most reflective of tension. The hypothesis under investigation here is that different states of underlying need-tensions are reflected in differential picture selection patterns. Guertin's study invokes a hidden assumption of the existence of the appropriate need-tensions, since there is no external criteria of such tensions. Further, there is no indication of the extent of homogeneity or heterogeneity of the group in regard to the variables under consideration. Therefore, the only justifiable conclusions would seem to be that hospital attendants can be characterized by certain negative and unbalanced category selections. Whether these characteristics are peculiar to the attendant population is unknown.

Fosberg compared category "like" and "dislike" patterns for a normal and a patient group with chance probabilities.[4] The results indicated that both groups differed significantly from chance selection but not *from each other.* The conclusion is drawn that the psychological factors at work in the test behavior are not

---

[4] Fosberg's miscalculation of probabilities has been pointed out by Cohen in a recent note (8).

those hypothesized by Deri and Szondi. As with the Guertin study, the existence of certain need-tensions within the subjects is assumed without any supportive evidence. Therefore, the results of a comparison with chance are difficult to interpret. On the other hand, the test's inability to differentiate between patient and nonclinical groups would seem to discredit the Szondi hypotheses. However, Fosberg does not specify how many or which categories should hypothetically distinguish the groups on the basis of single profiles. The main hypothesis of pathology regarding fluctuations in certain categories has not been considered. Thus, normal and pathological groups might evidence significant differences in stability for a series of profiles without any single profile distinctions.

Hill compared category, vector, and individual picture selection patterns of 1,066 college students with chance probabilities of such patterns. He found that most choice patterns differed significantly from chance. However, he points out that failure of chance to account for the test phenomena does not provide any information as to what psychological forces are actually operative. Deri's concept of need-tension remains neither verified nor rejected. Hill did find that males and females differed significantly in selections of sexual (homosexual and sadistic categories) and cyclic (depressive and manic categories) pictures, but the source of these differences is not known.

The most extensive normative data reported to date are in Harrower's (23) report of single profiles for some 1300 normals and neurotics. She records the distributions of the various selection patterns for each of several designated subgroups. Harrower merely states results without any discussion of pertinent Szondi hypotheses.

Deri makes repeated reference to certain syndromes of category selection with the implication of correlation of choices between the pertinent categories. Borstelmann (4) analyzed single profiles of 400 college students using a group administration, and found no evidence of covariation between any pairs of categories hypothesized as having a consistent relationship. Thus, the occurrence of a given combination of categories is no greater than would be expected by chance. There would seem to be no justification for the inference of common elements among certain categories. Borstelmann's results further indicated no sexual differences in choice patterns; a predominantly positive selection of paranoid and manic pictures; a predominantly negative reaction to hysteric, catatonic, and depressive pictures; and more variable selection of homosexual, sadist, and epileptic pictures. Although some support was evident for Deri's characterization of a young adult population, the choice patterns in the paranoid and depressive categories were opposite from those hypothesized.

Cole and Roberts (10) gave a group administration of the test to 50 students of introductory psychology at ten consecutive class sessions. Most of the pictures elicited a general positive or negative, rather than variable, reaction from the group. There was no consistency in group reaction among category members. There was some evidence of sexual differences and support for Deri's statements of trends. There is unfortunate confusion here between group and individual trends, as well as between stimulus and response consistency. Cole has not provided the crucial analysis of his data, the consistency

or fluctuation of group and/or individual responses to the test materials. Consideration of the data as 500 single profiles tends to obscure the behavior most meaningful to Deri's hypotheses regarding temporal effects.

Helme (24) did a factor analysis of category intercorrelations for single profiles of 72 young adults. Category selections were expressed in terms of "directive scores" (algebraic sum of plus and minus choices, regardless of loading). A simple cluster analysis resulted in "a major cluster for both males and females, comprising the sexual (homosexual and sadist) and ego (catatonic and paranoid) categories and the hysteric category attached to the sexual categories" (p. 5). This and other clusters are interpreted in accordance with Deri's hypotheses. Note that Helme, in contrast to Borstelmann, found both category intercorrelation and sexual differences. The contradictory evidence may be a function of the method of analysis. Borstelmann used a chi-square test of covariation for the four main scores of Deri (plus, minus, ambivalent, open). Helme's algebraic sums offer finer distinctions of category selection but at the same time combine in a single score several of the patterns used by Deri. This extended scoring schema may offer potentially more meaning for test interpretation, but it is not appropriate to Deri's hypotheses.

Finally Spitz (38) briefly reports on some of the results from a study of 110 Swiss kindergarten children. Although she notes that the children's choices are strongly affected by such peripheral factors as beards, the validity of dynamic sources for picture selection are taken for granted. Thus, test results differing from Szondi's results with Hungarian children are interpreted as personality-cultural variations. There is no apparent recognition that the stimulus properties of the pictures may differ for the two groups.

In summary, the evidence is rather clear that chance does not account for Szondi Test behavior but the nature of the psychological forces at work remains somewhat obscure. The accumulated results of test behavior for various groups are difficult to appraise owing to a lack of specification of the relationships between the test variables and known characteristics of the groups. Use of the test hypotheses to explain the test results does not provide critical evidence for the validity of these hypotheses. The method of empirical groups as a test of hypotheses would seem to have value only to the extent that the groups are clearly differentiated in regard to pertinent characteristics. Unspecified overlap between the groups will tend to obscure the relationships between variables. Selection of subjects who have known characteristics clearly relevant to specified test variables would seem to offer a more definitive approach to the investigation of interpretive hypotheses.

Stability and variability of test behavior. As previously indicated, the traditional correlational concept of psychometric consistency is not applicable to phenomena involving expectations of change or variability. Deri hypothesizes that the extent and location of change in test behavior are reflections of intrasubject variability.

Cole (9) compared initial and subsequent protocols for 86 college students. Seventy-four subjects had change of direction for two to five of the eight test categories. The changes were mostly of a minor nature, with few actual reversals of direction within a category. No cate-

gory was more changeable or stable than the others. Cole merely concludes that a single Szondi profile is not reliable. However, these results tend to support the proposition that the test does not reflect such relatively stable functions as cognitive processes and is, therefore, not reliable in the traditional sense of the word. The important question is whether the observed changes in test behavior are meaningfully reflective of subject functions or merely fortuitous.

Deri hypothesizes that fluctuations in test behavior are due to intrasubject instability, greater variability being generally more indicative of pathology. David and Rabinowitz (13) gave a series of six individual administrations to 20 paranoid schizophrenics and 20 student nurses. A Szondi Instability Score was computed on the basis of successive reactions (like, dislike, ignore) to individual pictures. The patient group had significantly greater variability, but not for any specific category.

Although this study represents a limited sampling, it seems that the results tend to support the general hypothesis of the relationship between pathology and instability of test performance. However, the hypothesis of greater instability of schizophrenics in the ego vector (catatonic and paranoid categories) is not borne out. The greater variability of psychotics' test behavior may be attributed to a general instability of functioning rather than to the specific dynamics hypothesized by Deri. David's method of analyzing reaction changes to specific pictures is particularly profitable because it avoids the unwarranted assumption of intracategory equivalence of stimuli. Thus, test change may be appraised in terms of any stimulus groupings that have demonstrable significance. It would be in-

teresting to analyze the results further. Is the group difference due to more variable reactions to certain pictures? Is there any consistency of reaction to pictures within existing categories?

The relation of test change to change in overt behavior has been carefully studied by Paine (30). Six test profiles given at two-day intervals to 30 functional psychotics were compared with daily ratings by attendants on a 15-item behavior scale. Test change was expressed as differences in choices for each category, divided into five gradients of change comparable to Deri's classes. There were no relationships between behavior and test changes. Subjects with the greatest test change were not consistently more unstable or severe in pathology. Open reactions were not related to overt behavior. The open-ambivalent ratio for the test series was not indicative of either pathology or control. Finally, the extent and kinds of change were not different from an independent group of college students.

These completely negative results must be due to either inappropriate study design or inadequacies of basic hypotheses. If one argues that the overt behavior observed herein is not relevant to the personality characteristics reflected in the test, then one is bound to specify more adequately the phenotypical referents of the hypothesized variables. Meanwhile, the results must be interpreted as casting some doubt on existing hypotheses about the test.

An interesting approach to a systematic investigation of unreliability has been suggested by Holt's (26) study of intra-individual covariation for a single subject in twelve administrations of the Szondi Test, Murray's "Mind-Reading" test, and self-ratings on a 130-item question-

naire. Numerous "significant" correlations (.70 and greater) between fluctuations are reported. The pertinence of such methodology to the intricate problem of reliability and validity in personality measurement is discussed by Holt. However, in the absence of more precise specification of the relation of test behavior to the total functioning of the individual, evidence of covariation among assessment techniques will constitute only presumptive validity of the test hypotheses.

Cahill (6) infers from Deri's exposition a hypothesis of greater intra-individual complexity and hence more variable test behavior for individuals of higher intelligence. He found no differences between high school students of low average and superior IQ's. Test change (five administrations) was expressed in terms of any change of choice patterns as a percentage of total possible change with a given category. This method of measuring change in test behavior allows for an element not controlled by other procedures, the actual freedom for alteration which is not equivalent for all patterns (plus, open, ambivalent, etc.). However, the basic hypothesis of the study seems to be a misinterpretation of Deri's position regarding level of intellectual ability. Differences in intelligence lend interpretative qualification to the hypotheses but *are not* postulated as actually producing differences in test patterns. The results actually support Deri's proposition that test behavior is a function of something other than cognitive processes.

In summary, variability seems to be a normal, nonclinical function of Szondi Test behavior. Though not extreme, this variability seems to be pervasive and not differential among test categories. There is greater, though still undifferentiated, variability among pathological subjects. Whether the changes in test behavior are related to hypothesized variables of intra-individual functioning is unknown; existing evidence is somewhat contradictory and indefinite. Clarification and verification of interpretive hypotheses regarding stability or instability of response will require extensive systematic study of intra-individual variability in terms of carefully specified relationships between test behavior and other functioning of the individual.

Experimental change. The studies discussed in the preceding section were concerned with the normal or usual fluctuations in test performance as reflections of intra-individual variability. Further clarification of the meaningfulness of variability or stability in test behavior may be elicited by use of the experimental situation method. The essential question is the kind and extent of test change that will result from the interpolation of specified independent variables.

Scherer (36) studied the effects of administration procedure by testing 22 mental patients with a battery of instruments, including the Szondi, in both individual and group situations. Certain significant differences in Szondi profiles were obtained and interpreted in terms of Deri's hypotheses. In this study, the lack of a priori specification of expected differences necessitates a posteriori speculation as to the psychological significance of the obtained results. Such a procedure necessarily involves the questionable presumption that the hypotheses are valid. The results can therefore be considered only as suggestive for further research.

Another study by Scherer (37) was

designed to assess the effects of the examiner upon subject behavior. Five paranoid schizophrenics were tested by five different examiners on successive days, and each examiner self-administered the test immediately following the subject's test. Five similar subjects were tested for five days with a group administration and asked to match each picture to one of eight descriptive phrases representing the test categories. The results showed that group profiles differed from individual administration only in greater plus choices for paranoid pictures. There was some influence (unspecified) of the examiner's personality. No significant daily variation was found in either patients or examiners. Some unique as well as common individual picture patterns for patients and examiners were evidenced.

This ambitious study design is an excellent example of attempting to do too much with too little. Examiner effects upon test behavior is an important area for study, and variation of examiners would seem to offer an appropriate method. But use of the test itself as a measure of the examiner is rather like pulling one's self up by one's own bootstrap. The presumption that the examiner-subject interaction will be reflected in the instrument being studied is gratuitous. Experimentation requires investigation of, not defense of, the hypotheses being studied. Scherer's approach should be encouraged under conditions of external criteria for both subjects and examiners.

Saunders and North (34) have tested the hypothesis that test behavior should not be sensitive to transitory sets if the behavior is due to basic personality structure. Four hundred and twenty students were divided into three groups with each receiving two standard administrations and an interpolated test of altered instructions. In the interpolated tests Group A was given the standard instructions of like-dislike selection, Group B was instructed to make happy-unhappy choices, and Group C was instructed to make strong versus weak character selections. Three of the six tests sets were used for the pre- and posttests, with the remaining three used for the interpolated test. The extent of agreement between pre- and posttest selections was essentially equivalent for all three groups, a mean of 75 per cent duplicate reactions. Thus, the test behavior does not seem to be affected by a transitory alteration of set toward supposedly equivalent stimulus materials. However, it should be noted that the assumption of equivalent stimulus values from set to set has not been upheld. In the limited sense that test-retest agreement exceeds chance expectations, the test may be said to be reliably assessing whatever forces are operative in the test behavior.

Odes's study of the effects of exposure to a supposedly strong emotional stimulus (29) is indicative of the inadequacies of experimental control that tend to confuse investigation of projective techniques. Forty graduate students were given two administrations in a single session with interpolated reading of either a short horror or humorous story. A statement is made by Odes that "a majority of the critical subjects had strong emotional reactions to the (horror) story and gave notable changes in their Szondi profiles after the story" (p. 3). But there is no indication of the criterion of "strong emotional reactions." Apparently there was no attempt to control contamination between assessment of the independent and de-

provide sufficient scope to reflect individuality. The Szondi Test requires expression of affective attitudes toward stimuli which are readily identifiable (as contrasted with the Rorschach or even the Thematic Apperception Test). Individuality is presumed to be expressed within the limits of a forced-choice task, freedom of individual response being restricted to affective selection among limited alternatives of recognizable stimuli.

3. Affective selection is assumed to be based upon certain generally associated meanings for given classes of stimuli. A given category of pictures is assumed to have equivalent associative values, yet potentially divergent affective meanings. The patterns of affective selection purportedly reflect individuality in terms of intra-individual complexity rather than as syndromes of deviation from normative patterns.

4. Finally, there is the assumption that behavior in one modality may be inferred from that in another modality. The nonverbal Szondi behavior is presumed to be predictive of the verbal behavior of symbolization and even such motoric behavior as expressive movement.

Three frames of reference have been advanced to explain the phenomena of affective selections among the Szondi pictures. Szondi himself maintained that the eight clinical syndromes pictured in the test represent certain personality characteristics which are hereditarily determined and manifest in physical appearance. Subjects are assumed to react to the test pictures on the basis of their own patterns of recessive or dominant genes. The knowledge of gene structure derived in this way leads to inferences about the subject's personality.

Deri, omitting the genetic assumptions of Szondi, hypothesizes that the clinical syndromes of the test are extreme manifestations of personality characteristics present in all persons. The affective selections by the subject are due to unconscious identification with or rejection of these characteristics within himself.

Both Szondi's and Deri's formulations assume that the dynamics of a given syndrome are reflected in the test pictures and reacted to as such by the subjects. However, only some of the test pictures have been found to have the stimulus values appropriate to such rationale.

The present writers have maintained that the mediational assumptions implicit in the Szondi and Deri formulations are not essential to the use of the test. A more empirical frame of reference in regard to the test phenomena is proposed. Like any other projective technique, the Szondi Test may be considered as consisting of partially structured stimulus materials which the subject is asked to handle in his own personal manner, thereby revealing something of his unique personality processes. The subject expresses affective attitudes towards the pictures in a direct manner by selection as like, dislike, or relatively neutral. These affective choices would seem to be based upon implicit and unverbalized associations to the stimuli. Inferences about individual characteristics would thus seem to be justified if the following necessary and sufficient assumptions maintain: (a) Subjects tend to assign consistent, though unverbalized associations to the pictures. (b) In selecting among the pictures, the subject responds in terms of personal affective, rather than stereotyped cultural, values. (c) Reaction tendencies toward pictures of common associative

values have interpretive significance which can be demonstrated by external criteria.

Within this empirical frame of reference as differentiated from Szondi and Deri, the test pictures may be restructured in accordance with demonstrated stimulus values. Any tenuous assumptions about manifest physical expression of personality are therefore unnecessary. With pictures of stimulus values appropriate to the test rationale, hypotheses relating to particular reaction patterns may be definitively investigated. Without a demonstrable rationale, the validity of the interpretive hypotheses will remain uncertain.

The Validity of Interpretative Hypotheses

The body of interpretive hypotheses for such techniques as the Rorschach, Thematic Apperception and Szondi tests are designed to assess individuality of functioning as reflected in the pattern of responses to a standard set of stimuli. The methods of analysis or evaluation (scores, abstracted themas, etc.) provide a means for ordering the unified personal productions of the subject in terms of the interpreter's conceptual and experiential framework. In this sense the "gestalt" of the subject's performance is somewhat arbitrarily destroyed and distorted in terms of the interpreter's orientation. Thus, the separate hypotheses formulated do not represent distinct and independent bits of response to be combined in a purely summative manner. Each hypothesis must have a core of generalized inference which is valid for all subjects so responding, yet capable of qualification in terms of relationships to other hypotheses. Here lies the seat of much confusion about validation of isolated hypotheses. Only the generalized hypothesis can be investigated by group differences. The qualifications, in terms of related hypotheses, require increasingly idiographic verification. Behavior must be studied in the context in which it is exhibited, not as it suits the convenience of the investigator.

An illustration of investigation of the validity of generalized inference from a single aspect of test-phenomena is provided in an unpublished study of Bardsley, Borstelmann, and Klopfer (2). The hypothesis under investigation was the inference of differential handling of passive needs from different reaction patterns to "homosexual" pictures of the Szondi test. The selection of stimuli as adequate for representation of such needs was based upon the Klopfer-Borstelmann study of associative stimulus values (27). The stimulus pictures were presented in rearranged groups so that each "homosexual" picture had relatively equivalent potential for selection (see Borstelmann-Klopfer [5]). Subjects were given a standard administration of the restructured Szondi test, as well as various other techniques and interviews specifically designed to elicit evidence of the functioning of passive needs. The results then provided a basis for ascertaining whether differential responses in particular Szondi behavior were similarly reflected in external criteria of the postulated variable. The point intended to be stressed here is not the end result of verification or refutation of the particular hypothesis, but the adequacy of methodology for validation of any interpretive hypothesis. The present writers maintain that any body of interpretive hypotheses must be capable of systematic verification.

A plea for scientific rigor in the validation of projective techniques inevitably raises the ghost of "clinical intuition." Some clarification of this factor in the interpretive process seems appropriate. The present writers maintain that intuition is not a function peculiar to the clinician. Rather, intuition would seem to refer to subliminal and implicit hypotheses that the investigator (clinical or experimental) has acquired in the course of repeated exposure to given phenomena. In so far as clinical intuition refers to recognition of the complex, interactive nature of individual behavior, it serves an important function. As a defense against explicit systematic investigation, it seems completely untenable. The progress of knowledge is dependent upon adequate communication. What seems to be called for is a more explicit operational statement of the implicit hypotheses used by the clinician. Systematic investigation of the clinician in the process of interpretation should be encouraged.

The question naturally arises as to whether such a technique as the Szondi Test is worthy of the rigorous investigations essential to determination of validity. Certain unique contributions of such a technique should be noted. The very structure of the forced-choice task which restricts the freedom of individual expression also yields a desirable objectivity of administration and scoring. Although interpretation remains a tenuous process of undetermined validity, the data forming the basis of interpretive inference can be established with a high degree of consistency among observers. The simplicity of the task required of the subject permits a wide range of application, from children to old age, feebleminded to superior mentality, normal to severe pathology, etc. Finally, the technique is unique in the use of a temporal series of behavior samples. Greatest interpretive emphasis is placed upon indices of change or stability. As such the test seems peculiarly well suited for the study of personality processes. If the Szondi Test could be put on a sound theoretical and empirical footing, it would seem to be a valuable addition to clinical and research tools.

Perhaps indicative of a favorable prognosis for projective techniques is the absence of studies dealing solely with the ability of the Szondi Test to discriminate among diagnostic groups. The inadequacies of nosological groups as external criteria for instruments that are *not* designed primarily to make such group differentiations have been repeatedly stressed by recent writers. Yet the literature concerning the Rorschach and other projective techniques abounds with studies of this kind. The present writers are encouraged to believe that research in the area of clinical assessment has begun to evidence sophistication in regard to methodology and study design pertinent to investigation of the assumptions underlying multidimensional techniques. To be sure, the present review has found the existing status of research with such an instrument as the Szondi Test far from being a paragon of experimental virtue. Yet the focus upon the adequacy of the rationale involved in the nature of the Szondi Test is indicative of a trend toward more fundamental and definitive inquiry. Any clinical instrument is not a shibboleth to be defended or refuted, but a provocative tool which requires careful investigation as to the nature of its contribution to the understanding of individual behavior.

REFERENCES

1. AINSWORTH, MARY D. Some problems of validation of projective techniques. *Brit. J. med. Psychol.*, 1951, **24**, 151–161.

2. BARDSLEY, R., BORSTELMANN, L. J., & KLOPFER, W. G. Expression of passive needs in the Szondi Test. Unpublished study, Univer. of California, 1950.

3. BLESSING, H. D., BEDFORD, G. S., & GLAD, D. D. An experimental investigation of some incidental features of the Szondi Test. *J. Colo.-Wyo. Acad. Sci.*, 1950, **4**, 65. (Abstract)

4. BORSTELMANN, L. J. Szondi picture selection patterns of 400 college students. Paper read at Midwest. Psychol. Ass., 1951.

5. BORSTELMANN, L. J., & KLOPFER, W. G. Does the Szondi Test reflect individuality? *J. Pers.*, 1951, **19**, 421–439.

6. CAHILL, R. F. The role of intelligence in changes within the Szondi Test profiles. *J. clin. Psychol.*, 1951, **7**, 379–381.

7. COHEN, J. The chance distribution of Szondi valences. *J. consult. Psychol.*, 1951, **15**, 130–133.

8. COHEN, J. A note on Fosberg's "Four experiments with the Szondi Test." *J. consult. Psychol.*, 1951, **15**, 511.

9. COLE, D. The reliability of a single Szondi profile. *J. clin. Psychol.*, 1951, **7**, 383–384.

10. COLE, D., & ROBERTS, E. Szondi results in group testing with college students. *J. clin. Psychol.*, 1950, **6**, 381–386.

11. DAVID, H. P. An inquiry into the Szondi pictures. *J. abnorm. soc. Psychol.*, 1950, **45**, 735–737.

12. DAVID, H. P. Szondi bibliography. *Psychol. Serv. Center J.*, 1952, in press.

13. DAVID, H. P., & RABINOWITZ, W. The development of a Szondi instability score. *J. consult. Psychol.*, 1951, **15**, 334–336.

14. DAVIDSON, W. N., MURPHY, M. M., & NEWTON, B. W. Experimental analysis of the Szondi Test. *Amer. Psychologist*, 1949, **4**, 388. (Abstract)

15. DERI, SUSAN. *Introduction to the Szondi Test, theory and practice.* New York: Grune and Stratton, 1949.

16. DERI, SUSAN. The Szondi Test: its application in a research study of depressive patients before and after electric-shock treatment. In L. E. Abt and L. Bellak (Eds.), *Projective psychology.* New York: Alfred Knopf, 1950.

17. FOSBERG, I. A. Four experiments with the Szondi Test. *J. consult. Psychol.*, 1951, **15**, 39–44.

18. GUERTIN, W. H. A consideration of factor loadings on the Szondi Test. *J. clin. Psychol.*, 1950, **6**, 262–266.

19. GUERTIN, W. H. A test of a basic assumption of the Szondi Test. *J. consult. Psychol.*, 1950, **14**, 404–407.

20. GUERTIN, W. H. A comparison of the stimulus values of Szondi's pictures with those of North Americans. *J. clin. Psychol.*, 1951, **7**, 163–166.

21. GUERTIN, W. H. A factor analysis of some Szondi pictures. *J. clin. Psychol.*, 1951, **7**, 232–235.

22. GUERTIN, W. H., & McMAHAN, H. G. A survey of Szondi research. *Amer. J. Psychiat.*, 1951, **108**, 180–184.

23. HARROWER, MOLLY R. Experimental studies with the Szondi Test. *Szondi Newsltr*, 1949, **1**, Suppl.

24. HELME, W. H. Study of intercorrelations of Szondi factors. *Szondi Newsltr*, 1950, **2**, 5–6.

25. HILL, V. T. The Szondi Test and chance. *Szondi Newsltr*, 1951, **3**, 1–16.

26. HOLT, R. R. An approach to the validation of the Szondi Test through a systematic study of unreliability. *J. proj. Tech.*, 1950, **14**, 435–444.

27. KLOPFER, W. G., & BORSTELMANN, L. J. The associative valences of the Szondi pictures. *J. Pers.*, 1950, **19**, 172–188.

28. LUBIN, A., & MALLOY, M. An empirical test of some assumptions underlying the Szondi Test. *J. abnorm. soc. Psychol.*, 1951, **46**, 480–484.

29. ODES, ZENIA. A study of experimentally induced changes in responses to the Szondi Test. *Szondi Newsltr*, 1950, **2**, 1–14.

30. PAINE, H. Association of measurable changes in Szondi Test profiles with measurable factors in behavior of psychotics. Paper read at Midwest. Psychol. Ass., 1950.

31. RABIN, A. I. Szondi's pictures: identification of diagnoses. *J. abnorm. soc. Psychol.*, 1950, **45**, 392–395.

32. RABIN, A. I. Szondi's pictures: effects of formal training on ability to identify diagnoses. *J. consult. Psychol.*, 1950, **14**, 400–403.

33. ROSENZWEIG, S. A method of validation by successive clinical predictions. *J. abnorm. soc. Psychol.*, 1950, **45**, 507–510.

34. SAUNDERS, R. E., & NORTH, A. J. The ef-

fect of an experimentally established frame of reference on the consistency of responses on the Szondi Test. *Szondi Newsltr*, 1951, 3, 1–7.

35. SCHAFER, R. Review of Deri's "Introduction to the Szondi Test." *J. abnorm. soc. Psychol.*, 1950, 45, 184–188.

36. SCHERER, I. W. The psychological scores of mental patients in an individual and group testing situation. *J. clin. Psychol.*, 1949, 5, 405–408.

37. SCHERER, I. W., WINNE, J. F., PAGE, H. A., & LIPTON, H. An analysis of patient-examiner interaction with the Szondi pictures. *J. proj. Tech.*, 1951, 15, 419–420. (Abstract)

38. SPITZ, CHARLOTTE. Szondi Test and age; experimental studies of children from 5 to 7 years. *Szondi Newsltr*, 1950, 2, 1–4.

39. SZOLLOSI, E., LAMPHIEAR, D. E., & BEST, H. L. The stimulus values of the Szondi pictures. *J. consult. Psychol.*, 1951, 15, 419–424.

40. SZONDI, L. *Szondi Test; experimentelle Triebdiagnostik.* Bern: Hans Huber, 1947.

41. SZONDI, L. *Experimental diagnostics of drives.* New York: Grune and Stratton, 1952.

42. SZONDI, L. *Szondi Test; experimentelle Triebdiagnostik; Testband.* Bern: Hans Huber, 1947.

43. SZONDI, L. *Schicksalsanalyse.* (Rev. Ed.) Basel: Benno Schwabe, 1948.

44. WALLEN, R. Factors affecting the choice of certain Szondi Test pictures. *J. consult. Psychol.*, 1951, 15, 210–215.

45. *Szondi Newsletter.* Ed. by Dr. Virgil T. Hill, mimeographed at the University of Oklahoma, Norman, Okla.

SOME RELATIONS BETWEEN TWO STATISTICAL APPROACHES TO ACCIDENT PRONENESS[1]

WILSE B. WEBB   AND   EDWARD R. JONES
Washington University   *Human Resources Research Laboratories*

In the pursuit of the elusive concept of "accident proneness" a number of statistical techniques have been used. Recently, several articles on the relative merits of two of the more widely used techniques, the correlational method and the "Poisson fit" method, have appeared (1, 2, 5, 6). The present article is about an apparently basic identity between these methods. From our analysis it appears that a tempest may be brewing in the teapot.

The "Poisson fit" technique was developed by Newbold in 1927 (7) and by Cobb, independently, in 1940 (3). This method is an extension of the method of simply comparing an obtained distribution of accident occurrences of indivuduals with a Poisson distribution. Such a comparison can indicate the presence of "other than chance" factors but gives limited information about the extent of the "other than chance" factors. The Newbold and Cobb attempt to correct this deficiency is based upon the fact that the mean and the variance of a Poisson distribution are identical. Under certain assumptions, the difference between the mean and the variance of an obtained accident distribution may be taken as representative of the degree of accident proneness in the population from

which the distribution was obtained. As stated by Thorndike:

If we have a series of single distributions, each of which is a Poisson, and wish to combine them into a single overall distribution, the mean of the total distribution will be the weighted average of the means of the sub-distributions . . . for each sub-group, the variance around its own mean is equal to that mean. The weighted average of the within-groups variance will, therefore, equal the weighted average of the sub-group means, i.e., the mean for the total group. Having this estimate of the within-groups variance, one can compute the total variance for the empirical distribution, and then one can immediately estimate the between-groups variance by subtraction (8, p. 34).

These obtained variances may, in turn, be converted into a more useful form by noting that the between groups variance divided by the total variance is equal to R^2. This statistic in a sense represents the reliability coefficient of the accident criterion. The square root of this result is equivalent to the index of reliability.

Assuming, then, that the subdistributions of our population are Poissons, this seems a legitimate method of estimating the extent to which "other than chance" factors are operative in our population. This method is that preferred by Blum and Mintz (1).

The other method that we are concerned with is the correlational method. Its assumptions are straightforward enough. The period in which the subjects are operating is divided into two half periods. It is

assumed that if a differential consistency of accident behavior is operative in the subjects this will result in a correlation between the two periods. That is, the highly prone individual in one period will be highly prone in the other period, the individuals of low proneness will exhibit this characteristic in the two periods, and a correlation of the accident behavior of the individuals between the two periods will reveal the extent of the consistency of accident performance above and beyond factors distributed independently of the individual's performance. This method is preferred by Maritz (5).

The two methods described above were applied to an Air Force population in an attempt to determine the degree of "other than chance" factors operative in the Air Force flying situation.[2] The "Poisson fit" method was applied to the distribution given in Table 1. The resultant R^2 was .204.

into accidents occurring on odd days and even day. Fortunately, this was easily accomplished by IBM methods since the cards used contained the day, the month, and the year of occurrence for any given accident.

The resultant correlational table is given in Fig. 1. The product-moment correlation obtained was .107. Corrected by the Spearman-Brown formula, this correlation was .193.

In working with this table it was noted that the diagonal frequencies of the table approximated a binomial distribution for each diagonal. Further consideration led to the conclusion that a binomial expansion of the obtained accident frequencies would represent the best theoretical distribution of these accidents in a bivariate division. It was reasoned as follows: given one accident, the probability of the accident occurring in one or the other of two independently determined periods would be equal.

TABLE 1
DISTRIBUTION OF INDIVIDUAL NONFATAL ACCIDENTS DURING A PERIOD OF AIR FORCE FLYING

Number of Accidents	0	1	2	3	4	5	6	7
Number of Pilots	12475	4117	1016	269	53	14	6	2

The correlational method was also applied to this population. Because of changing assignments and varying circumstances of flying between periods of Air Force flying, the periods were not divided by a first-half–last-half method. Rather, in an attempt to equalize experience, exposure, etc., the data were divided

Given a number of single accidents and two "periods," the most likely distribution of the accidents would be that of half the accidents occurring in one period and half the accidents in the other. Given a sample of individuals having two accidents with equal probability of occurrence in either of two periods, the most likely distribution of these accidents would be the occurrence of one-fourth of both the accidents occurring in one period; one-fourth of both the accidents occurring in the other period; and half the accidents distrib-

[2] The actual time period and circumstances of operation for the population have been purposely omitted for security purposes. This also fortunately limits inferences concerning Air Force accident proneness, which deserves a more detailed statement.

	0	1	2	3	4	5	6
6	0	1	0	0	0	0	0
5	0	1	0	0	0	0	0
4	4	4	1	1	0	0	0
3	29	10	5	2	0	0	0
2	234	101	26	4	1	0	0
1	2045	513	100	7	2	0	0
0	12475	2071	269	39	6	0	1

Accidents Occurring on Odd Days—Even Years—Even Days—Odd Years (vertical axis)

Accidents Occurring on Odd Days—Odd Years—Evens Days—Even Years (horizontal axis)

FIG. 1. DISTRIBUTION BY PERIOD OF OCCURRENCE OF INDIVIDUAL NONFATAL ACCIDENTS

uting themselves so that one accident occurred in one period and one in the other period. This distribution, of course, is that obtained by a binomial expansion with a .50 probability, taking into account the number of events.

It then was hypothesized that accidents will distribute themselves binomially between two periods if the accidents are independent of the method of division of the periods. Any other distribution would indicate that the method of division was related to the accident performance of the subjects.

Following this assumption, we constructed a correlational table from the originally obtained accident distribution given in Table 1. This

distribution is given in Fig. 2. It will be noted that each diagonal is a binomial distribution (the 1-1 coordinates, the 2-2 coordinates, etc.). The obtained correlation was .112. An application of the Spearman-Brown formula resulted in an r of .201. It is interesting to note that this form of distribution is precisely that which was generated by Burke in a separate article on this problem utilizing different assumptions (2).

This correlation looked familiar. It was in fact identical to the second decimal place with the R^2 obtained by the analysis of variance method. To check the possibility that this was a strange coincidence rather than a factual similarity of results for the two methods, a number of other dis-

	0	1	2	3	4	5
5	0.4	0.0	0.0	0.0	0.0	0.0
4	3.3	2.0	1.0	1.0	0.0	0.0
3	33.6	13.2	4.0	2.0	1.0	0.0
2	254.0	101.0	20.0	4.0	1.0	0.0
1	2058.5	508.0	101.0	13.2	2.0	0.0
0	12475.0	2058.5	254.0	33.6	3.3	0.4

FIG. 2. BINOMIAL EXPANSION OF ACCIDENT DISTRIBUTION

tributions were obtained and the two methods were applied. The results in these cases were identical to the third decimal place.

A mathematical analysis of the bivariate binomially derived distribution was performed by Burke. A note by him appears following this article. Essentially his analysis indicates the mathematical identity of the two methods.

Certain conclusions may be drawn:

1. Two methods of estimating the theoretical extent of "accident proneness" may be derived from essentially independent assumptions. One, the analysis of variance technique, makes its assumptions concerning the sub-distributions of the obtained data. The other, the binomial correlational method, assumes a chance distribution of accidents for any subject between two periods of operation.

2. Operationally, the two methods yield identical estimates.

3. Mathematically, the identity of the methods may be shown.

4. Theoretically, many of the arguments concerning the relative merits of the two methods must become suspect.

5. Practically, the results indicate that the choice of the two methods outlined becomes dependent only upon convenience, ease of conceptualization, or personal preference.

REFERENCES

1. BLUM, M. L., & MINTZ, A. Correlation versus curve fitting in research on accident proneness: reply to Maritz. *Psychol. Bull.*, 1951, 48, 413–418.
2. BURKE, C. J. A chi-square test for "proneness" in accident data. *Psychol. Bull.*, 1951, 48, 496–504.
3. COBB, P. W. The limit of usefulness of accident rate as a measure of accident proneness. *J. appl. Psychol.*, 1940, 24, 154–159.
4. CRAMÉR, H. *Mathematical methods of statistics.* Princeton: Princeton Univer. Press, 1946.
5. MARITZ, J. S. On the validity of inferences shown from fitting of Poisson and negative binomial distributions to observed accident data. *Psychol. Bull.*, 1950, 47, 434–443.
6. MINTZ, A., & BLUM, M. L. An examination of the accident proneness concept. *J. appl. Psychol.*, 1949, 33, 195–211.
7. NEWBOLD, E. M. Practical application of the statistics of repeated events, particularly of industrial accidents. *J. roy stat. Soc.*, 1927, 90, 487.
8. THORNDIKE, R. L. The human factor in accidents. USAF Sch. Aviat. Med., Project No. 21-30-001, Report No. 1, *Restricted.*

NOTES CONCERNING THE WEBB-JONES ARTICLE

C. J. BURKE

Indiana University

We consider a set of accident data on N individuals. To each individual there corresponds a number z, the number of accidents he has had during the entire observational period. The period is divided into two exclusive and exhaustive periods of equal length; for each individual, x and y represent the number of accidents he has had in each of these periods. Clearly, x, y, and z are discrete random variables and

(1)
$$z = x + y.$$

We associate the following distributions with the variables:

(2) p_i. (the probability that $x = i$)

(3) $p_{.j}$ (the probability that $y = j$)

(4) p_k (the probability that $z = k$)

(5) p_{ij} (the probability that $x = i$ and $y = j$).

The "binomially-partitioned" distribution introduced by Webb and Jones (5) is obtained in the following way. The expected number of individuals who have precisely k accidents during the entire period is given by:

(6)
$$n_k = np_k. \qquad (4)$$

For each k, these n_k values of $z = x + y$ are partitioned into specific values of x and y according to the succes-

sive terms in the expansion of $(\frac{1}{2} + \frac{1}{2})^k$. Hence, we have:

(7)
$$p_{ij} = p_k \binom{k}{i} \left(\frac{1}{2}\right)^k$$
$$= p_{i+j} \binom{i+j}{i} \left(\frac{1}{2}\right)^{i+j}. \qquad (1)$$

Where $\binom{k}{i}$ is the number of combinations of k objects taken i at a time.

We wish to establish two important facts about the distribution defined by (7).

I. The correlation between x and y, corrected by the Spearman-Brown formula, yields the expression used by Newbold (4) and by Cobb (2).

For the average value of z, we have:

(8)
$$\overline{Z} = \sum_k k p_k. \qquad (4)$$

For the average value of z^2, we have:

(9)
$$\overline{Z^2} = \sum_k k^2 p_k. \qquad (4)$$

For the variance of z, we have the well-known expression:

(10) $\sigma^2 = \overline{Z^2} - \overline{Z}^2$. Cramér (3, p. 180).

For the various average values with respect to x and y, always equal by symmetry, we have:

(11)
$$\bar{x} = \bar{y} = \sum_k p_k \sum_{i=0}^k i \binom{k}{i} \left(\frac{1}{2}\right)^k \qquad (5), (7)$$
$$= \sum_k p_k \frac{k}{2}$$
$$= \frac{\overline{Z}}{2} \qquad (8)$$

$$(12) \qquad \bar{x}^2 = \bar{y}^2 = \sum_k p_k \sum_{i=0}^{k} i^2 \binom{k}{i} \left(\frac{1}{2}\right)^k \qquad (5),\ (7)$$

$$= \sum_k p_k \frac{k^2 + k}{4}$$

$$= \frac{\overline{Z^2} + \overline{Z}}{4} \qquad (8),\ (9)$$

$$(13) \qquad \overline{xy} = \sum_k p_k \sum_{i=0}^{k} i(k - i) \binom{k}{i} \left(\frac{1}{2}\right)^k \qquad (5),\ (7)$$

$$= \sum_k p_k \left(\frac{k^2}{2} - \frac{k^2 + k}{4}\right)$$

$$= \frac{\overline{Z^2} - \overline{Z}}{4} \qquad (8),\ (9)$$

$$(14) \qquad \sigma_x^2 = \sigma_y^2 = \overline{x^2} - \bar{x}^2$$

$$= \frac{\sigma^2 - \overline{Z}}{4}. \qquad (10),\ (11),\ (12)$$

For the correlation coefficient between x and y, we have:

$$(15) \quad r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y}$$

$$= \frac{\sigma^2 - \overline{Z}}{\sigma^2 + \overline{Z}}. \quad (10),\ (11),\ (13),\ (14)$$

The Spearman-Brown formula used by Webb and Jones is:

$$(16) \quad r \text{ corrected} = \frac{2r}{1 + r}$$

$$= \frac{\sigma^2 - \overline{Z}}{\sigma^2} \qquad (15)$$

the expression used by Newbold and by Cobb.

II. We say that x and y are independent if and only if

$$(17) \qquad p_{ij} = p_{i.} \cdot p_{.j}.$$

We wish to show that, if z has a Poisson distribution, the distribution obtained by the binomial partitioning is identical with the distribution obtained from Poisson marginal distributions under the hypothesis of independence.

If z has a Poisson distribution so that

$$(18) \quad p_k = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{Cramér } (3, \text{ p. } 203)$$

we obtain for the joint distribution of x and y by binomial partitioning:

$$(19) \quad p_{ij} = \frac{\lambda^k}{k!} e^{-\lambda} \binom{k}{i} \left(\frac{1}{2}\right)^k \quad (7),\ (18)$$

$$= \frac{\left(\dfrac{\lambda}{2}\right)^{i+j}}{i!\,j!}\, e^{-\lambda}. \qquad (1)$$

$$(22) \quad p_{ij} = \frac{\left(\dfrac{\lambda}{2}\right)^{i+j}}{i!\,j!}\, e^{-\lambda}. \quad (17),\ (20),\ (21)$$

Assuming identical Poisson distributions for x and y, we write:

$$(20) \qquad p_{i.} = \frac{\left(\dfrac{\lambda}{2}\right)^{i}}{i!}\, e^{-\lambda/2}$$

$$(21) \qquad p_{.j} = \frac{\left(\dfrac{\lambda}{2}\right)^{j}}{j!}\, e^{-\lambda/2}.$$

Under the hypothesis of independence, we obtain:

Since (19) and (22) are identical, II has been established. This means that the binomially partitioned distribution introduced by Webb and Jones is identical with one of the cases discussed by the writer (1) in a chi-square procedure. It should be noted that in the special case represented by (18) (Cramér [3, p. 204]),

$$(23) \qquad \sigma^2 = \overline{Z} = \lambda \qquad (8),\ (9),\ (10)$$

and

$$(24) \quad r \text{ corrected} = r = 0. \quad (15),\ (16),\ (23)$$

REFERENCES

1. Burke, C. J. A chi-square test for "proneness" in accident data. *Psychol. Bull.*, 1951, 48, 496–504.

2. Cobb, P. W. The limit of usefulness of accident rate as a measure of accident proneness. *J. appl. Psychol.*, 1940, 24, 154–159.

3. Cramér, H. *Mathematical methods of statistics.* Princeton: Princeton Univer. Press, 1946.

4. Newbold, E. M. Practical application of the statistics of repeated events, particularly of industrial accidents. *J. roy. stat. Soc.*, 1927, 90, 487.

5. Webb, W. B., & Jones, E. R. Some relations between two statistical approaches to accident proneness. *Psychol. Bull.*, 1953, 50, 133–136.

PATTERN ANALYSIS: THE CONFIGURAL APPROACH
TO PREDICTIVE MEASUREMENT[1]

EUGENE L. GAIER AND MARILYN C. LEE

University of Illinois

One of the more promising trends in current psychometric research is an increasing concern with methods of evaluating patterns of test scores and test responses. In clinical, vocational, social, and educational psychology, there is a growing consensus that taking account of interrelationships among test variables will contribute appreciably to the efficiency of prediction. Evidence is being offered that higher validities may be obtained if predictive variables are treated as patterns rather than as mere "and-summations" or averages of separate and independent scores.

The purpose of the present paper is to review current theoretical and technical developments in pattern analysis with particular reference to its most recent application, the study of item response patterns within a single test. In this latter instance, our initial hypothesis is that consideration of response configurations will yield more fruitful results than the usual method of reporting merely the total score for a test. Zubin (28) pointed out that total scores may conceal as much as they reveal, since the psychological equivalence of the constituent items can seldom be established nor their additive character proved. A total score may thus carry considerably less diagnostic significance than a direct and detailed analysis of the test responses *per se*.

In the present review, attention will be devoted first to considerations emerging from the study of total score patterns. Subsequently, the application of these principles to the analysis of response patterns within a single test will be examined. Our basic assumption is that test data may be so treated as to yield a higher degree of predictive utility than that obtainable by the more traditional additive methods.

ASPECTS OF PATTERN ANALYSIS

The problem of pattern analysis has at least three aspects which should be differentiated for purposes of conceptual clarity. These may be designated as (a) *pattern representation*, (b) *pattern matching*, and (c) *pattern prediction*.

The adequate representation of a pattern in numerical terms is in itself a problem which necessitates the prior resolution of several statistical difficulties. The main problem is one of developing indices which can simultaneously depict both the amounts or elevations of the several variables composing the pattern and the interrelationships or overlap among them. Another problem is that of correcting for the errors of measurement characteristic of all psychometric instruments, in order to accurately estimate the "true" profile. This becomes a question of preventing the effects of unreliability from contaminating the statistical

representations of the profiles studied. Such problems must be solved before one can proceed to the second aspect of pattern analysis—that of pattern matching.

In order to determine the degree of congruence between two patterns, it is again necessary to develop indices which can simultaneously consider both the level and the configuration[2] of the variables employed. The subsequent discussion of attempts to solve this problem will serve to illuminate the inadequacies of existing methods as well as some of the basic theoretical and methodological considerations involved.

The third facet of pattern analysis involves the prediction of complex criteria by the simultaneous use of several variables. This aspect of pattern analytic methodology shares certain features in common with both pattern matching and traditional multiple prediction problems. If the criterion to be predicted is itself formulated in terms of a profile, as is the case in many clinical and sociological studies, then the predictive problem becomes one of ascertaining the degree of similarity between the undiagnosed (predictive) profile and the reference (criterial) profile. However, if the criterion is a single composite index, as is frequently the case in vocational and educational measurement, then the investigator's problem involves ascertaining that amount and configuration of variables which will yield the best prediction of the global criterion involved. *Pattern prediction* may thus include *pattern matching*, and both may depend heavily upon the prior development of adequate methods of *pattern representation*.

[2] Level refers to the amount or intensity of any given trait. Configuration takes into consideration the interrelatedness of the variables.

These distinctions underlie the subsequent discussion of pattern analytic theory and methodology. Any statistical technique which employs several variables and attempts to capitalize upon the predictive power contributed by a consideration of their interrelationships should find a place in the above classification.

ANALYZING A PATTERN

Cattell (2) has pointed out that since a pattern consists of both elements and relations, a duality exists in every pattern which must be considered in its definition. For predictive purposes, however, it is possible to divorce these two aspects of a psychometric profile and to study them either separately or in combination. Thus attention may be concentrated only upon the interrelationships among the elements, it may consider both configuration and level simultaneously, or it may be concerned merely with the levels or elevations of the component variables.

Exclusive concern with either shape or elevation alone ignores the possibility that these two characteristics of a pattern may be psychologically interdependent. Although some writers do not consider both level and configuration essential to maximal discrimination, it has been asserted by Du Mas (8), and recently emphasized by Osgood and Suci (20), that similar patterns located at different levels may indicate interpretations quite at variance with each other so far as behavior is concerned. As the latter authors point out, "Measurement specialists have often expressed concern, for example, over the fact that a moron and a genius may have closely correlated profiles, despite the absolute discrepancy between their scores" (20, p. 251).

It is thus apparent that although

psychological measurement still depends heavily on level alone for discrimination, there is a growing belief (2, 5, 8, 14, 19, 28) that (a) configuration is important in its own right, (b) it is more than a mere summation of constituent elements, and (c) it contributes most to prediction when taken in conjunction with level (i.e., with data on the relative amounts of the profile variables being studied).

These principles have received widespread application in recent vocational and educational research (11, 12, 25, 26). In the 1930's, an early attempt to objectify intuitive methods of interpreting aptitude test profiles was made by the Minnesota Employment Stabilization Research Institute (11, 21). Batteries of tests were administered to persons in a variety of occupations in an effort to ascertain the nature of their test-score patternings. The findings indicated that individuals successful in a given occupation tend to manifest patterns of ability quite distinct from the patterns exhibited by persons in other occupations and in the general population. Subsequent testing of individuals with the same batteries thus gave vocational guidance a more solid foundation by permitting the individual profiles to be compared for similarity with those previously found to characterize the various occupational groups. This rationale has directed the subsequent construction and use of such standard batteries as the Differential Aptitude Tests and the USES General Aptitude Test Battery. A significant aspect of these developments is that recognition is given to the importance of the configuration as well as to the amount of the separate aptitudes measured.

In clinical psychology, the analogues of pattern are the concepts of *type* and *syndrome*. Although *types* have fallen into disrepute because of their tendency to multiply and their lack of specificity regarding component variables (25), this concept might be revived if sufficiently powerful statistical techniques were available.[3]

The *syndrome*, no doubt because of its respectable medical parentage, has retained sufficient popularity to stimulate attempts at statistical refinement. The concept of the syndrome underlies the development of multiscore clinical tests which yield configural results based on several different tests or subtests.

Cronbach (5) has recently suggested that the data obtained from multiscore tests (e.g., Rorschach, MMPI, Wechsler-Bellevue) and factorial inventories (Guilford, Thursone) may perhaps best be conceived as points in k-dimensional space, where k equals the total number of traits measured. The study of patterns is then resolved into the attempt to describe how these multiple measurements are distributed in k space and how the distributions compare with one another. From this point of view, the problem of pattern analysis becomes that of identifying regions of score concentration in k space, and determining the significance of the obtained differences in the distances among those scores.

Although the statistical comparison of score profiles is logically justified by the k-space concept, Cronbach emphasizes that the existing methodology of such analysis is still inadequate. A primary difficulty is that of accurately determining the extent to which two patterns agree when the component scores of either or both may be unequally reliable and unequally correlated. Under such conditions, confidence limits for the significance of obtained differ-

[3] Stephenson's (21, 22) application of the Q-sort technique to Jungian typology has been an attempt in this direction.

ences are difficult to establish. Other problems as yet untouched or inadequately handled include those of curvilinear distributions and multimodal concentrations in k space, as well as that of dealing with the changes with time which may occur in any of the dimensions measured.

In summary, the shortcomings of current methods for dealing with the multiple discrimination problems involved in pattern matching include failure to correct for unfulfilled assumptions of (a) linearity, (b) equality of reliability, and (c) equality of intercorrelations among items or tests. In addition, some techniques fail to take account of the direction of obtained differences, or even more important, concentrate on configuration to the *exclusion* of level. However, their contributions to objectivity of interpretation justify brief mention of several representative approaches to this problem.

COMPARING TEST PATTERNS

Zubin's similarity coefficient (28). This is a relatively inexact method dealing only with dichotomous scores. It is based on the number of variables for which the scores are the same, divided by the total number of scores in the pattern. This technique ignores both elevation and the expected frequency of agreements (2).

Du Mas' "slope" method (7, 9, 10). This quick but rather gross method describes the similarity of profiles in terms of the slopes of corresponding segments. Though taking into account both level and configuration, the method is imprecise and can eventuate in maximally good prediction when the two profiles being compared are actually dissimilar (8).

The r_p coefficient of pattern similarity. Developed by Cattell (2), this coefficient yields an over-all index of profile relations based on the discrepancies between means for each

variable, and thus can be used as a measure of absolute agreement between two patterns. However, it may not take sufficient account of configuration in some cases, and is based on the assumption of uncorrelated and normally distributed components (2).

The D coefficient. Cronbach and Gleser (6) have developed an index (D) of dissimilarity between score sets based on the distances among score coordinates in k-dimensional space. Similar to a measure recently developed by Osgood and Suci (20), this coefficient has a known distribution and thus can be used for testing the significance of obtained differences between groups or individuals. It is useful even when the variables employed are intercorrelated, but in this case leads to problems of weighting, since all variables are not equally important or general.

The checklist method. Frequently utilized in the study of Rorschach protocols, this technique utilized lists of *signs* characteristic of some special group serving to differentiate it from other groups. The checklist total score essentially represents a multiplication of single-score methods and thus tends to ignore configuration and to lose information when the component variables are intercorrelated. It has the further disadvantage of capitalizing on differences due to accidents of sampling (5).

Ratios among the profile variables. The reliability of the variables is of particular importance in the use of this method, since minor score changes may drastically affect the size of the obtained ratios (4). Moreover, the technique ignores level and tends to grow cumbersome if the number of variables involved is large (2).

Correlation coefficients among the scores. The use of product-moment r's ignores the relative amounts of

the components being compared and for that reason is unable to yield a measure of absolute agreement (congruity) between two patterns (2). Although its singular assumption of linearity of relationship may be circumvented by the use of eta coefficients, the method in either case makes it possible to obtain perfect comparison when the two profiles are actually very dissimilar in both level and pattern (8). Q technique tends to restrict study to those subdivisions of *k*-dimensional space where many cases fall (5) and groups patterns according to gross similarities without adequate reference to level (2).

Pattern tabulation. Developed by Cronbach (3), this method employs normalized scores and profiles expressed in terms of the deviation of such scores from their average for each person or group. The technique can handle only two or three variables at a time, and functions best when the scores utilized are equally reliable and equally correlated (4). In using this procedure one must assume that profiles having the same shape are psychologically equivalent, since it ignores level by defining configuration in terms of deviation units.

Multiple regression equations. This technique considers the intercorrelations of scores and weights them appropriately, giving good predictions even when dealing with unreliable figures containing large random elements (2, 4). However, it may prove unstable on cross validation because random variation often falls to the experimenter's advantage by contributing maximally to the weights first obtained. Like the factorial specification equation, which is a more specific expression, multiple regression assumes linearity and takes no account of the possible optimum level of each of the variables (2). In addition, the composite index it yields is based on a combination of several variables and consequently throws away much information by reducing several dimensions to one (5).

Discriminant function. This formula will discriminate two categories of subjects from a mixed sample and thus is of value in studies comparing different types of subjects (4). Like the multiple regression equation, however, it loses information by reducing several dimensions to one composite index. Although it does correct for intercorrelation and unequal reliability of variables, problems of weighting tend to diminish the efficiency of this technique.

Many of the procedures so far developed for the description and comparison of score patterns suffer either from a lack of flexibility or from inability to meet the necessary statistical and psychological assumptions. As pointed out above, improved devices are needed for estimating true profiles, for weighting the factors appropriately and for evaluating the significance of obtained differences when samples are compared via orthogonal models on many unequally fallible, correlated measurements. Moreover, statistical treatment frequently requires that the data be reduced to fewer dimensions by considering certain patterns to be equivalent so that samples sufficiently large may be obtained to insure adequate reliability (5). But despite their deficiencies, the emphasis on patterning inherent in these techniques represents a much needed trend toward the development of objective means for representing and evaluating the organization of ability and personality traits as they occur both in the individual and the group.

ANALYSIS OF RESPONSE PATTERNS

The first explicit recognition of the predictive advantages to be gained from the study of items within a

single test appears to be attributable to Brigham (1). In his *Study of Error*, published in 1932, Brigham summed up the results of six years' work with the Scholastic Aptitude Test in the following statement:

These studies might be summarized by the general conclusion that a test item, regarded as a specific problem situation to which a certain number of answers may be made, when given to two or more populations sampled for study in the same manner, will show approximately the same distribution of criterion scores for each possible answer in the several groups sampled (1, p. 69).

Brigham regarded this finding as having great importance for educational research, since it demonstrated that important information may be lost if all persons who fail an item are assigned a score of zero for that item and are thereby treated alike. The results obtained suggested that it is possible to specify patterns of response peculiar to and characteristic of particular groups of people. Brigham thus emphasized the potential fruitfulness for predictive purposes of breaking up the typical datum of educational measurement (i.e., the total score in a series of tests or subtests) into smaller units consisting of particular test items. By studying individual test items as specific situations and by analyzing the patterning of errors, he hoped to isolate specific response configurations which could be used to differentiate various classes or types of individuals from each other and from mixed samples.

Study of the significance of associations among test items was subsequently taken up by Zubin (27, 28), who pointed out that the traditional practice of assigning empirical weights to responses and summing these to obtain a total score assumes unjustifiably that numerically equal scores are psychologically equivalent.

Zubin contended that a knowledge of the patterns of test responses may be more important than knowledge of the total number correctly passed. He hypothesized that response patterns, or clusters of responses to a given group of items having a frequency greater than that expected by chance, will yield results of greater diagnostic significance than a total score which represents merely the summation of discrete responses to qualitatively different items.

To test this hypothesis, Zubin selected a number of personality inventory items, each of which was believed capable of discriminating between normal and abnormal groups. Combining these items into triads, he tested for the significance of pattern differences by the dichotomous score method outlined above (*Zubin's similarity coefficient*). It was found that scoring by patterns rather than by individual responses is possible, and that normal subjects show significantly more consistency of pattern than psychiatric patients. The patients exhibited several unique patterns characteristic only of their own diagnostic group. There appeared to be differences among the psychiatric categories in the extent to which they displayed normal and unique patterns of response. Although each item that entered into any of the triads was itself diagnostic, a further significant finding was that not all of the triads of these individually discriminatory items proved to be diagnostic as a pattern.

The other side of this coin has recently been considered by Meehl (19), who points out that configural scoring of an item pair may give perfect discrimination between two groups even though each item taken separately has zero validity. Thus, dichotomous items having no discriminatory power by themselves might separate two population groups

completely when scored with respect to pattern.

Lubin[4] has appealed to Rao's (22) maximum likelihood scoring technique for a general solution to "Meehl's paradox." According to Lubin, Meehl's finding applies to continuous as well as to dichotomous variables, and to the case where both groups are scored on but one variable (differentiation being achieved by means of individual item scores). Rao's method applies to these and to all other problems which involve assigning subjects to one of several categories and computing the proportion of misclassifications to be found. Essentially, it involves reducing the number of misclassifications to a minimum by calculating the likelihood of a particular response pattern for each group and then assigning all subjects with this pattern to the group where the likelihood is maximum.

The studies reviewed suggest that configural analysis as applied to item responses within a single test may yield better discriminations than the more usual additive techniques which ignore inter-item relationships. It is implied in the rationale of such analysis that the joint presence of several responses carries a diagnostic significance not inherent in those responses when taken by themselves. Less explicitly, this concept is inherent in the practice of developing a number of different scoring keys for the same test or inventory on the basis of empirical data showing the relationships of these items to various criteria. The utilization of patterns of response represented by the development of such keys not only yields higher validity coefficients but also makes it possible to use the same psychometric instrument for a variety of predictive tasks (14).

These considerations underlie much of the research undertaken by McQuitty (15, 16, 17) since his attempt in 1937 to differentiate between normal and psychiatric groups by the use of response patterns. McQuitty's investigations constitute perhaps the most extensive attack made to date on the diagnostic significance of response patterns in personality questionnaires. The most recent development of his technique (18) involves ascertaining the degree of association between every response alternative and every other response alternative for all items in the test being studied. The obtained matrix of association coefficients is factor analyzed, and a table giving the factor loading of each response alternative is prepared. "Factor measurements" for a given individual may then be obtained from this table by noting the factor loadings for each of the responses chosen by that individual. From such loading patterns, it becomes possible to ascertain the relative degree of integration manifest in the person's responses.[5]

Thus, if the subject has chosen many response alternatives characteristic of a large percentage of the original group, i.e., many responses having high loadings on the first factor, his pattern may be said to represent majority opinion and his personality said to be integrated according to contemporary cultural standards. If, on the other hand, his responses may be accounted for only by means of a large number of factors, the subject may be regarded as relatively disintegrated, since his response pattern suggests the presence of attributes characteristic of several disparate groups. This interpretation

---

[4] LUBIN, A. Maximum likelihood scoring: a solution to Meehl's paradox. Unpublished manuscript.

[5] McQUITTY, L. L. Implications of certain measures of personality integration for theories of social psychology. Unpublished manuscript.

is supported by other evidence indicating that group conformity is highly related to personality integration in the sense that an inventory of items weighted to measure the latter discriminates between community persons and mental hospital patients (18).

The basic hypothesis tested in these studies was that personality is integrated to the extent that the individual's ideas are consistent with each other, and subsequently to the extent that his responses to a personality inventory may be accounted for in terms of a single factor. Through its attempt to heighten discriminatory power by taking into account the interrelations among test items, this method of "attribute analysis" underlines the potentialities inherent in the configural approach to the study of test data. Moreover, its usefulness is not restricted to personality test materials. As applied to ability tests, the significant aspect of the technique is that the relation of each test response, whether correct or incorrect, is computed with reference to every other response.[6] In contrast to the usual total-score approach, the method therefore provides for a new interpretation of wrong responses, on the rationale that the individual does not respond randomly when he does not know the correct answer. That there may be meaningfulness in the individual's pattern of wrong responses or choice of plausible distractors is an hypothesis which attribute-analysis is eminently suited to test.

SUMMARY

The studies reviewed indicate that considerable predictive advantage is to be gained by giving adequate statistical recognition to the configural aspects of test scores and responses. Research to date has demonstrated that the discriminatory power of a prognostic instrument is partially a function of how it is scored, and that an analysis which takes into account patterns of response may have more predictive usefulness than the traditional total-score approach. The questions at issue here are but a special case of one of psychology's general problems, i.e., that of describing the relative frequencies of response in each of the classes it has abstracted from the behavior complex. Recent theory has questioned the utility of traditional additive techniques and has focused attention on the analysis of patterns and dependent probabilities among both tests and test responses. The trend of the evidence suggests that this new direction of emphasis will in the future be to the decided advantage of the predictive endeavor.

[6] Gaier, Lee, and McQuitty (13) have recently utilized attribute-analysis in an attempt to isolate patterns of reasoning from response interrelationships characterizing a test of logical inference. Five interpretable factors, tentatively identified as distinct response sets or patterns of response, emerged from a 120-variable factor analysis (i.e., intercorrelations of 24 items, each with five possible answers). The findings suggest that attempts to capitalize upon the consistency manifest in response patterns may considerably augment the amount of information obtainable from predictive instruments.

REFERENCES

1. BRIGHAM, C. C. *A study of error.* New York: College Entrance Examination Board, 1932.

2. CATTELL, R. B. r_p and other coefficients of pattern similarity. *Psychometrika,* 1949, **14**, 279–298.

3. CRONBACH, L. J. "Pattern tabulation": A statistical method for analysis of limited patterns of scores, with particular reference to the Rorschach test. *Educ. psychol. Measmt,* 1949, **9**, 149–171.

4. CRONBACH, L. J. Statistical methods applied to Rorschach scores: A review. *Psychol. Bull.*, 1949, **46**, 393–429.

5. CRONBACH, L. J. Statistical methods for multiscore tests. *J. clin. Psychol.*, 1950, **6**, 21–25.

6. CRONBACH, L. J., & GLESER, GOLDINE C. Similarity between persons and related problems of profile analysis. Technical Report No. 2, Bureau of Research and Service, University of Illinois, April, 1952. (Mimeo.)

7. DU MAS, F. M. A quick method of analyzing the similarity of profiles. *J. clin. Psychol.*, 1946, **2**, 80–83.

8. DU MAS, F. M. On the interpretation of personality profiles. *J. clin. Psychol.*, 1947, **3**, 57–65.

9. DU MAS, F. M. The coefficient of profile similarity. *J. clin. Psychol.*, 1949, **5**, 123–131.

10. DU MAS, F. M. A note on the coefficient of profile similarity. *J. clin. Psychol.*, 1950, **3**, 300–301.

11. DVORAK, BEATRICE J. Differential occupational ability patterns. *Empl. Stab. Res. Inst. Bull.*, 1935, **3**, No. 8. Minneapolis: University of Minnesota Press.

12. DVORAK, BEATRICE J. The new U.S.E.S. general aptitude test battery. *J. appl. Psychol.*, 1947, **31**, 373–376.

13. GAIER, E. L., LEE, MARILYN C., & McQUITTY, L. L. Response patterns in a test of logical inference. Memorandum Report A-2, Training Research Laboratory, University of Illinois, October, 1952. (Mimeo.)

14. LECZNAR, W. B. Evaluation of a new technique for keying biographical inventories empirically. *Research Bulletin* 51-2, Human Resources Research Center, Lackland AFB, San Antonio, Texas, March, 1951.

15. McQUITTY, L. L. An approach to the measurement of individual differences in personality. *Charact. & Pers.*, 1938, **7**, 81–95.

16. McQUITTY, L. L. A measure of individual differences in personality integration. *Canad. J. Psychol.*, 1950, **4**, 171–178.

17. McQUITTY, L. L. Clinical implications for a measure of mental health. *J. abnorm. soc. Psychol.*, 1951, **46**, 73–78.

18. McQUITTY, L. L. A statistical method for studying personality integration. In O. H. Mowrer (Ed.), *Psychotherapy: A symposium on theory and research.* New York: Ronald, in press.

19. MEEHL, P. E. Configural scoring. *J. consult. Psychol.*, 1950, **14**, 165–171.

20. OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.

21. PATERSON, D. G., & DARLEY, J. G. *Men, women and jobs.* Minneapolis: Univer. of Minnesota Press, 1936.

22. RAO, C. R. Utilization of multiple measurements in problems of biological classifications. *J. roy. statist. Soc.,* Series B, 1948, **10**, 159–193.

23. STEPHENSON, W. Methodological consideration of Jung's typology. *J. ment. Sci.*, 1939, **85**, 185–205.

24. STEPHENSON, W. A statistical approach to typology: The study of trait universes. *J. clin. Psychol.*, 1950, **6**, 26–38.

25. TOOPS, H., & KUDER, G. F. Test construction and statistical interpretation. *Rev. educ. Res.*, 1935, **5**, 229–441.

26. TRABUE, M. R. Occupational ability patterns. *Personnel J.*, 1933, **11**, 344–351.

27. ZUBIN, J. A technique for pattern analysis. *Psychol. Bull.*, 1936, **33**, 773. (Abstract)

28. ZUBIN, J. The determination of response patterns in personality adjustment inventories. *J. educ. Psychol.*, 1937, **28**, 401–413.

As an experimental innovation, it has been decided to evaluate newly published textbooks in a given area by grouping them instead of following the traditional procedure of having a separate review for each individual book. Such area reviews will appear as SPECIAL REVIEWS. The review of *Some Recent Textbooks in Social Psychology* by M. Brewster Smith in this issue is the first attempt in this direction.

This issue also includes an example of multiple reviewing. Victor's book on handwriting is reviewed both by Werner Wolff and by Julian Rotter. It was felt that since some psychologists would react negatively and some would react positively to this book, an attempt should be made to have different kinds of reaction presented. In a controversial field, perhaps fair treatment can best be assured if more than one point of view is represented. It is planned that area reviews and multiple reviews will be given further trials.

W. D.
E. G.

SPECIAL REVIEW
SOME RECENT TEXTS IN SOCIAL PSYCHOLOGY

M. BREWSTER SMITH
New York City

ASCH, SOLOMON E. *Social psychology.* New York: Prentice-Hall, 1952. Pp. xvi+646. $5.50.

DOOB, LEONARD W. *Social psychology. An analysis of human behavior.* New York: Holt, 1952. Pp. xix+583. $5.00.

FARIS, ROBERT E. L. *Social psychology.* New York: Ronald, 1952. Pp. vii+420. $5.00.

HARTLEY, EUGENE L., & HARTLEY, RUTH E. *Fundamentals of social psychology.* New York: Knopf, 1952. Pp. xix+740. $5.50.

SWANSON, GUY E., NEWCOMB, THEODORE M., & HARTLEY, EUGENE L. *Readings in social psychology.* (Rev. Ed.) New York: Holt, 1952. Pp. xix+680. $5.00.

Social psychology, as its composite name suggests, is a marginal field. It is well known that the first two texts to bear the title came forth in the same year from a psychologist, McDougall (6), and a sociologist, Ross (10). The dialectic interplay between the psychological and sociological approaches that runs as a persistent theme through the subsequent development of the discipline is still with us, as the current crop of texts attests.

Over the years, social psychology seems rather to have suffered the handicaps of marginality than to have reaped its benefits: the characteristic difficulty of the marginal in establishing self-identity has been balanced only recently if at all by the creative ferment that marginality sometimes releases. Perhaps, it might be argued, the trouble was that the parent fields were too immature at the time they conceived social psychology to pose the right problems for their hopeful offspring to resolve: the "developmental tasks" may have been wrongly selected and paced. Or the methodological poverty of its family background may have left the infant science ill-equipped to make its way in the world of complex affairs that everyone agreed was its proper domain. At all events, the rapid growth and transformation of social psychology has come only in the recent postwar years. All of the books under review here bear the impress of these recent developments, though together they come to no resolution on the identity problem and they vary in the extent to which they give promise of emerging synthesis.

A further characteristic of social psychology was foreshadowed by Ross and McDougall: the landmarks in the development of the field have been its textbooks. It is doubtless a sign of persisting immaturity that the basic textbook with its necessarily limited scope still recommends itself to competent social psychologists as a vehicle for their major theoretical contributions. Yet most of the authors of the present texts have ambitions beyond the production of useful adjuncts to teaching, and one of them—Asch—has written a work of major theoretical significance.

Directions in Contemporary Texts: A Bird's-Eye View

It is the lack of consensus on the

boundaries and organizing conceptions of social psychology, of course, that challenges the textbook writer and makes him influential. This lack of consensus is amply reflected in our four 1952 texts, which virtually box the compass of current perspectives on the field. They are not, to be sure, located at cardinal points on pure dimensions, being pitched at different levels and blended with varying amounts of eclecticism. Taken together, however, they afford a good opportunity to get one's bearings on social psychology, not only as it is being taught to the present generation of undergraduates but also as it is giving direction to the ever-increasing stream of research activity.

Faris represents one direction in almost pure type: the line of interactionist social psychology out of Mead and Cooley. Faris is a sociologist, and to find a psychologist in the lineal ancestry of this sociological tradition one must go back to John Dewey and the birth of functionalism. Faris does not neglect the work of psychologists since Dewey, but neither does he draw much that is consequential from them. Mostly he views with critical alarm; besides that, he selects from the psychological literature to embroider and document his account. Another recent text in the same tradition is that by Lindesmith and Strauss (5).

Between them, the Hartleys and Doob fill in the region in which the center of gravity of psychological social psychology surely lies. The Hartleys represent the wing that has been more exposed to sociological thinking of the sort drawn on by Faris. For them, communication is the central social-psychological process and socialization the principal developmental focus. Role, status, and social norms provide the con-

ceptual armory for analyzing the individual's contemporary relations with groups. As for their assumptions about underlying psychological processes, they are eclectic and not particularly rigorous, drawing on learning research with no commitment about the nature of the learning process, placing considerable emphasis on perception, and making free use of the amalgam of psychoanalytic and anthropological thinking that has wide currency in discussions of culture and personality. The reader will recognize that this book belongs to a thriving genus of which recent texts by Newcomb (8) and Sargent (11) are also members.

Of the books under consideration, Doob's makes the least pretension to theoretical synthesis. Aimed squarely at the college sophomore, it shares with Britt's still recent book (2) a wide-ranging concern with interesting social phenomena without much emphasis on abstract theory. The perspective, all the same, is one of S-R reinforcement doctrine. While Doob wrestles with the claims of individualistic vs. group-centered approaches—a dichotomy that leads him into difficulties from which the other three interaction-centered texts are relatively free—there is little doubt that his sympathies lie in the individualistic direction. Since he disavows systematic objectives, one must still turn to Miller and Dollard (7) or to F. H. Allport's classic (1) for clear examples of American behaviorism in social psychology.

But it is Asch who has provided us with another classic to set against F. H. Allport. Writing from the position of Gestalt theory, Asch takes a stand far enough from the center of gravity to wield an Archimedean lever on the world of contemporary social psychology. While some of his strictures against the dominant psy-

chological trends resemble those made by Faris from his sociological bailiwick, Asch's analysis cuts much more deeply and shows surer discrimination. His closely reasoned examination of underlying psychological assumptions warrants the closest attention from the majority of American psychologists who will find themselves in frequent disagreement.

As for its positive emphases, the book does not attempt systematic coverage, however systematic its perspective. The Gestalt emphasis on cognitive processes informs his original treatment of interaction, and throws into prominence a variety of unorthodox problems on which his own considerable research contributions are presented among others. Of recent texts in social psychology, that by Krech and Crutchfield (4) resembles Asch most closely in orientation. Krech and Crutchfield write the more systematic text, but Asch's work is more searching and far-reaching in its implications.

Now for a closer look at these four texts, and at the revised book of readings by Swanson, Newcomb, and Hartley that samples so well the research literature on which they draw.

Faris

So far as its likely use in departments of psychology is concerned, the Faris book could be summarily dismissed. It deserves to be taken more seriously, however, since it is written by a competent sociologist and will undoubtedly be used in many of the social psychology courses taught in sociology departments.

The very first sentence establishes the tone of controversy against "individual" psychology that pervades the book:

The importance of social psychology

in contrast to individual psychology lies in the fact that the complex behavior of daily life is not fully explained by an inventory of the physiological system of man (p. 3).

It is the scarcely tacit assumption that the "individual" psychologists of psychology departments are mere psychophysiologists, who have left it to the sociological social psychologist to develop the only psychology with human relevance. It follows, then, that the author's task is to teach all the psychology his students will need. Better that they should not have had courses in general psychology, for if they have, there is much to unlearn. Accordingly, over half of the book is devoted to his own general introduction to psychology, under such chapter headings as "Inadequacy of Biological Motivation," "The Emergence of Consciousness," "The Concept of the Unconscious Mind," "Social Determination of Learning, Perception, and Memory," and "The Social Factor in Ability."

In keeping with his definition of the field of social psychology as "the organization and control of behavior in primary interaction" (p. 4), he then turns to primary group processes and role differentiation and conflict. Aside from a final cursory chapter on "Trends and Problems in Social Psychology," the remaining chapters on "The Neurotic Role," and "Personality Disorganization" would normally be treated in other courses by psychologists.

Since so much of the book is devoted to rewriting general psychology, the content of Faris' revision warrants attention. While there is a liberal sampling of citations from recent psychological literature, one's confidence in his judgment is weakened by the extent to which he leans

on secondary sources, particularly the standard elementary text edited by Boring, Langfeld, and Weld, from which he repeatedly quotes excerpts under the scarcely disarming phrase, "As Boring writes . . . " (see, for example, p. 88). Then, too, one who has graded tests in elementary psychology shivers a little to read of the Rohrschach (sic) test (p. 386).

There is no space to detail the many points in which psychologists are likely to regard his treatment as less than satisfactory. Let me simply list a few:

1. He espouses an extreme environmentalism, even in respect to the sources of exceptional abilities (p. 234).

2. His criticism of conditioning takes stimulus-substitution for granted, and completely by-passes Hull to attack old Pavlov (pp. 78–85).

3. In his extended exposition of a functionalist interpretation of consciousness he presents the still controversial motor theory as established fact (p. 95).

4. In his attack on physiological drives as a basis of motivation, he ignores the conventional distinction between drive and motive, and hence bests a straw man (pp. 18–23).

5. He spends a chapter on an irrelevant refutation of the "unconscious mind," in the course of which he rejects the concept of repression on such flimsy and ill-conceived grounds as these: repression is unnecessary for the explanation of forgetting (p. 122); subjects with hypnotic or hysterical anaesthesias can be tricked into making discriminations within the supposedly anaesthetized range (pp. 132–139); the associative imagery in Coleridge's Kubla Khan can be traced to sources in his previous reading (pp. 141–144).

6. His interpretation of the neurotic role solely in terms of secondary gain leads him to embrace what is implicitly a smug moralism. Thus, he speaks of a combat psychiatric case in these terms: "When he was brought wounded to the hospital he woke up shaking and nervous

—a condition which could be maintained as a *comfortable* method of escape from further combat and mental conflict" (p. 320, italics mine).

7. He introduces the sociogram as "a device somewhat similar to a scale" (p. 402).

Apart from his inadequate command of contemporary psychology, and his inadequate criticism of its vulnerable points, there is of course great merit in the interactionist position as he expounds it. And it is true enough that psychologists have too often given lip-service to the interactive sources of human nature in society without carrying out the radical revision of their concepts that such a position entails. It is also unfortunately true, however, that the interactionist theory which burst full-blown from the pens of Cooley and Mead, having generated singularly little research, remains in Faris' text essentially as its brilliant authors left it. Surely the most careful and ingenious research will be required to verify and refine the notions of role taking and empathy and self that remain fixed points in this sociological current. But the perfunctory roll-call of research techniques in his final chapter leaves Faris firmly entrenched in the discursive rather than experimental tradition.

The Hartleys

Instructors who want a fully packed and up-to-date text that acquaints the student with current thinking and research in the areas most frequently regarded as social psychological will probably prefer this book to the others in the present list. It is a fat book, replete not only with references to the experimental literature but with clinical case summaries and anthropological accounts. And it covers much ground judicious-

ly. The chapter on attitudes to which Clyde Hart has contributed is probably the best brief statement available on current research and theory on the topic.

Yet for all these strong points, the Hartleys' text is not fully satisfactory. It is loosely organized, the chapters being divided into many brief sections among which it is not hard to lose the thread. The major parts of the book, on communication, socialization, and group processes, are only very tenuously bound together, in spite of some repetitiveness. The treatment of underlying psychological processes is rather loose and uncritical. And the considerable number of misprints and mistakes in bibliographical citations bespeaks hasty editing. Again I shall list my more specific criticisms for the sake of brevity:

1. The essay by Gerhart Wiebe on "Mass Communications," included by the Hartleys in keeping with their practice of concluding each major part with an applied chapter, is special pleading. Without questioning Wiebe's data or the soundness of his arguments, one notes that he *is* arguing in defense of the commercial media, rather than examining dispassionately just what their impact may be.

2. "Self-interest" is proposed as the principal determinant of set (p. 42), though even without the benefit of Asch's trenchant criticism this hoary and ambiguous mainstay of popular psychologies would seem patently inadequate to the complexities of human motivation. To break it down into "learning and yearning," as the authors do later (p. 133), does not help.

3. Part II, on socialization, does not treat the learning of roles, a topic relegated to the third part. While this allocation may make for convenience of exposition, it is at the cost of theoretical integration.

4. The treatment of social norms suffers from the usual ambiguity between mere *uniformities* of social behavior and standards of *evaluation* that carry the compulsion of legitimacy. The chapter on "The Functioning of Social Norms" shows particularly poor organization, with dangling sections on intergroup relations and group productivity the relevance of which is unclear. In the latter section Thelen's list of principles for "change agents" is simply appended without discussion.

5. Examples are continually drawn from the field of prejudice and intergroup relations, one to which the Hartleys have of course made a notable contribution. But the net impact on the reader may convey a rather specialized notion of the preoccupations of social psychology.

6. In drawing eclectically on many sources, the authors sometimes sacrifice theoretical rigor. They adopt, for example, a stimulus-substitution diagram of the deconditioning of children's fears (p. 248). They reprint the list of devices promulgated by the Institute of Propaganda Analysis (p. 85) almost without discussion, in spite of the dubious psychological value of this classification, and the still more dubious prospect of communicating much to the student from a mere list.

7. Occasionally the authors' interpretations of research data seem bizarre to me. Thus the increase in intelligence test scores of Negro children after migration to the North is attributed to group pressures toward conformity (p. 427). For another example, the fact that children conceive of their parents as embodying multiple roles successively rather than simultaneously is interpreted as a sign of their restricted experience and intellectual limitations (p. 527), when the artificial abstractness of the assumed adult conception would seem the more remarkable.

The book remains a useful one, though it would be a better text if it were less inclusive, and more substantial theoretically if it were more tightly woven.

Doob

It is hardly fair to criticize Doob's book on points of theory, since it is so clearly designed as a text for a quite elementary course. In its own terms it succeeds very well indeed. Its marked virtues follow from Doob's restraint in setting himself to produce a textbook, not a research manual or a new theoretical synthesis. As a novel feature, a large number of brief and well-edited excerpts from original sources are interwoven in the text—descriptions of experiments, writings of classical theorists, sociological descriptions, even historical accounts and literary passages. These are admirably selected to attract student interest and provoke student thought. If the book is light on abstraction and systematization, it shows more than usual pedagogical skill in inculcating scientific modes of thinking about social behavior.

While I would therefore rank the book high among those of its level of difficulty and degree of popular interest, I must take serious issue with the underlying assumptions that structure his approach to the field. In a summary passage he states:

The orientation to the group can be on the level of the individual: other people may be viewed simply as additional stimuli. *But* it is so very difficult to apply those principles simultaneously to many individuals .... In addition, people in groups may behave so differently that a new set of principles on an entirely different level—the level of the group—seems necessary .... The protagonists in this controversy do not seem to be reaching agreement .... Maybe social psychology must function for the time being on both levels: let the group-orientation discover and pose new problems, and let the individual orientation attempt to solve them by supplying the

details and even the principles (pp. 554–555).

But the protagonists in the controversy are reaching more agreement than Doob believes—or, more accurately, they are gradually being replaced by others who have been reared to a definition of the situation that makes Doob's dilemma obsolete. It is unnecessary to remain trapped in the dull historic controversy of individualism vs. the Group Mind. Asch makes the most incisive analysis of how this false dichotomy arises from a failure to examine the psychological character of interaction. But even without Asch, the account of interaction that Hartley and other writers in the same vein have modified from the sociologists at least indicates the direction from which resolution is to be sought.

Doob's position has major consequences for the subject matter of his book. The relational variable of role, for example, is completely ignored, while social class, which can be regarded as either a group-level concept or a *property* of individuals, is given extensive treatment. Lacking the interactional conception, it seems fair to say, social psychology has no boundaries for Doob other than those of interest. There is no theoretical criterion for inclusion or exclusion. And, accordingly, one finds such miscellaneous items as a long—and educationally sound—treatment of the physical anthropology of race (pp. 161–183), and chapters on vogues, social conflicts, and the diffusion of change.

A word about the tone in which the book is cast. It seems to me both unnecessary and regrettable that Doob has often seen fit to fish for student sympathy and interest by self-deprecating remarks, flippancy, and the juicy tidbit. Doob's style is

otherwise sufficiently attractive and the content of enough intrinsic interest that these classroom tricks only detract from his otherwise laudable attempt to convey the scientific spirit.

Asch

If Doob writes for the sophomore, Asch writes for his professional colleagues, the graduate student, and advanced undergraduates who are ready to find excitement in a venture in critical and constructive thought. There are other books from which the student will learn more of the content of contemporary social psychology, but none that exposes the reader to so searching an examination of its assumptions and their consequences. A book that has obviously been years in the writing, it will be read for a long time to come. And its clear but often quietly eloquent style makes the reading a more than intellectual pleasure. Asch writes a thoroughly psychological social psychology that makes no claims to empire in the social sciences:

The study of social behavior is part of the task of general psychology. Its facts and principles cannot be derived from the study of behavior outside the social setting. At the same time a theory of society cannot be exclusively psychological. The interrelations among the actions of the members of society reveal regularities and tendencies that can be studied in their own right; these are the province of the social disciplines (p. 38).

If social psychology must build on its own distinctive facts and cannot perform its function either by the rote application of principles derived from asocial contexts or by beach-combing operations in other social sciences, as Asch's contention has it, it has not got very far on its task.

It has to be admitted that social psychol-

ogy today lives in the shadow of great doctrines of man that were formulated long before it appeared and that it has borrowed its leading ideas from neighboring regions of scientific thought and from the social philosophies of the modern period. It is paradoxical but true that social psychology has thus far made the least contribution to the questions that are its special concern and that it has as yet not significantly affected the conceptions it has borrowed (p. viii).

So Asch begins by examining some of these "doctrines of man," taking issue with contemporary themes in social psychology to which they give rise. He singles out the following as prevailing dogmas with which he is in disagreement:

The ego-centered character of men.
The supremacy of irrational emotions.
The primacy of rationalization in human thinking.
The basis of human experience in arbitrary association and conditioning.
The roots of adult attitudes in childhood experiences.

There is indeed no doubt that Asch is a deviant. Of the two theoretical currents that he identifies as most influential in social psychology, he rejects S-R reinforcement theory outright on standard Gestalt grounds. Psychoanalysis he finds wanting for its omissions and emphases rather than for being flatly wrong: "What is lacking in Freud's account," he says, "is the sense that society is the condition of freedom as well as a source of oppression" (p. 347). Oddly enough, he takes virtually no explicit notice of the additional current of sociological interactionist theorizing, though its principal contentions are fully embedded in his own thinking.

After an introductory exposition of the Gestalt position, he analyzes the problem of interaction from a cognitive standpoint. Would that

the lectures of G. H. Mead had been rendered with such lucidity!

The paramount fact about social interaction is that the participants stand on common ground, that they turn *toward one another*, that their acts interpenetrate and therefore regulate each other.... It is individuals with this particular capacity to turn toward one another who in concrete action validate and consolidate in each a *mutually shared field*, one that includes both the surroundings and one another's psychological properties as the objective sphere of action.... The interpenetration of viewpoints once and for all separates social action from acts of individuals which happen in combination to produce certain regular results. The process is not one in which individuals combine like gases to lose their identity and produce something different from either of them. Rather it requires that each participant retain his perspective and assert his individuality (pp. 160–162, italics his).

The most fundamental problem of social psychology, therefore, is the process by which veridical social perception is achieved. Unlike the Meadian theorists, Asch has of course himself tackled this interactional problem with considerable experimental ingenuity. The remainder of this section of the book is "An Introduction to Group Theory," in which, it seems to me, he effectively exorcises the spectres of extreme individualism and the group mind.

The ensuing section on social needs includes, among other things, a critical scrutiny of the doctrine of self-interest, a brief but provocative phenomenological analysis of rules and values, and a sophisticated critique of cultural relativism. The final section, "Effects of Group Conditions on Judgments and Attitudes," draws on his previously published examination of the doctrine of suggestion and contains a report on his recent experiments on the modifica-

tion of judgments by groups. A brief chapter then takes sympathetic but critical notice of research on small groups in the Lewinian tradition. Two chapters on opinions, attitudes, and sentiments take account of the more quantitative methods of investigation, but rightly insist that the phenomena need to be examined from a more psychological perspective. The final short chapter on propaganda examines recurrent features of exploitative propaganda, and emphasizes its limitations.

As a systematic examination of central problems of social psychology in the light of Gestalt theory, the book is an admirable performance. Its balance nevertheless suffers somewhat from the isolation of American Gestalt psychology and from Asch's almost religious dedication to rationality. Speaking of the reasons for a re-examination of psychological axioms, Asch writes that "there is also a human reason for undertaking [it]: if the grim picture psychology draws were correct, there would be no hope for man or society" (p. 30). Drawing the picture as sharply as he does, perhaps he is right. At all events his attempt to make tacit assumptions sharply explicit, if it does not always result in a fair or flattering portrait of non-Gestalt psychology, creates the conditions for theoretical advance. It demonstrates once and for all, moreover, the weakness of social psychologies that claim neutrality as to the nature of underlying psychological processes. But the fervor that motivates his criticism does make for a one-sided picture. In a scrupulous attempt to be just, Asch leaves the dark side of man's nature to psychoanalysis: he wisely abstains from trying to refute Freud on his own ground. The facts remain, however, that irrational proc-

esses exist, and that they remain residual to his system.

Since Asch makes no attempt to cover in full the conventional ground of social psychology, one cannot criticize him for obvious omissions. It is relevant, however, to point to a a major gap in his Gestalt doctrine of social interaction. It is a crucial assumption of his theory that bodily expressions of inner states mirror isomorphically the psychological processes that they accompany. Asch's brief discussion of this point not merely lacks convincing evidence, but fails to show how such postulated isomorphism can be given intelligible meaning.

Swanson, Newcomb, and Hartley

Since the book of readings edited by Newcomb and Hartley (9) has been used so widely during the past five years, the revised edition should receive a warm welcome. The number of separate pieces has been reduced from 81 to 65, of which between a third and a half are new with this edition. Simply to list a few of the newcomers highlights the far-reaching developments that make the initial edition out of date: Asch on the effects of group pressure, Leavitt on communication patterns, Festinger, Thibaut, and Back on interpersonal communication, Bales on interaction process analysis, McClelland and Friedman on cross-cultural research on need achievement, Merton and Kitt on reference group theory, Deutsch and Collins on interracial housing—the array is impressive. As before, some of the pieces were specially written or revised for this volume, a few are excerpted from books, but most are reprinted from the periodical literature. To this reviewer, the arrangement is an improvement over the earlier version, though in a collection of readings order is inconsequential.

In a field so active as social psychology has recently been, hardly anyone will quarrel with the editors' decision to draw heavily on new material. Individual opinion will differ on the wisdom of specific inclusions and omissions, but most social psychologists should applaud the total offering. The book provides ample supplementary reading for all of the texts reviewed here except Faris, and even for sociological courses of his pattern there are a number of directly relevant selections.

Instructors would do well to advise their libraries not to discard their battered copies of the original edition, as the half of the old volume that has not been reprinted contains much material that still is useful.

Some Comments and Speculations

Perhaps it is unduly rash to attempt any general comments on the heterogeneous company of texts we have passed in review. I should like to close, however, with two general remarks: one about the teaching of social psychology and the other about trends in theory.

The books reflect no consensus on the level at which social psychology should be taught or on how much general psychology should be presupposed. The common practice would seem to be to write into a social psychology text all the general psychology that the student presumably needs for the purpose at hand. Often this appears to be very little. In following this custom, our textbook writers may be bowing to realistic necessity, but few teachers would agree, I think, that the situation is satisfactory. Doubtless there will continue to be a legitimate demand for books at a variety of levels

of difficulty. It is of some interest to note, however, that Buxton, Cofer, and their collaborators (3) in their recommendations for a revised undergraduate curriculum in psychology suggest that social psychology be placed at the advanced level in their three-step hierarchy. If their proposals should be widely adopted, a different and better pattern of textbook writing would be required.

Despite the seeming theoretical chaos to which these books bear witness, I would venture that the old tug-of-war between sociological and psychological conceptions of the field is undergoing a change, one that will leave stranded those who are too fond of the old controversies. Three of the four texts center their theory around a conception of interaction. Once this is accepted in its full consequences, the causes for battle vanish, though there remain important special considerations for the psychologist and the sociologist who approach the common ground from different directions. As the recent paper by Sears (12) gives promise, reinforcement as well as cognitive theory can be given an interactionist phrasing. The alternative to this focus, and I think a mistaken one, is to resign oneself to confusion about the contents of a field with no organizing principle and no boundaries except those set by personal preference.

REFERENCES

1. ALLPORT, F. H. *Social psychology.* Boston: Houghton Mifflin, 1924.
2. BRITT, S. H. *Social psychology of modern life.* (Rev. Ed.) New York: Rinehart, 1949.
3. BUXTON, C. E., COFER, C. N., GUSTAD, J. W., MACLEOD, R. B., MCKEACHIE, W. J., & WOLFLE, D. *Improving undergraduate instruction in psychology.* New York: Macmillan, 1952.
4. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
5. LINDESMITH, A. R., & STRAUSS, A. L. *Social psychology.* New York: Dryden, 1949.
6. MCDOUGALL, W. *An introduction to social psychology.* London: Methuen, 1908.
7. MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
8. NEWCOMB, T. M. *Social psychology.* New York: Dryden, 1950.
9. NEWCOMB, T. M., & HARTLEY, E. L. *Readings in social psychology.* New York: Holt, 1947.
10. ROSS, E. A. *Social psychology.* New York: Macmillan, 1908.
11. SARGENT, S. S. *Social psychology.* New York: Ronald, 1950.
12. SEARS, R. R. A theoretical framework for personality and social behavior. *Amer. Psychologist,* 1951, 6, 476–483.

METTLER, F. A. (Ed.) (The Columbia Greystone Associates, Second Group.) *Psychosurgical problems.* New York: Blakiston, 1952. Pp. xii+357. $7.00.

This book is a sequel to the one produced by the original Columbia Greystone group, titled "Selective Partial Ablation of the Frontal Cortex," 1949 (cf. *Psychological Bulletin,* 1951, 48, 185 ff.). It has a number of avowed purposes: (*a*) to compare the less commonly used psychosurgical procedures involving cortical venous ligation, thermocoagulation, thalamotomy and transorbital lobotomy; (*b*) to explore the value of new psychometric tests in assessing clinical status; (*c*) to evaluate the biological effects of transorbital lobotomy; (*d*) to report the status of patients, studied in the first project, two years or more after their operation; and (*e*) to find out why psychosurgery was followed in the first project by recovery in 20 per cent more of the cases than for a control group. Similarly to the first project, the second one is the product of many hands and many different disciplines. The administrative burden of coordinating so diverse an effort must have been considerable.

This second study was beset by many problems. The criteria for the selection of patients were so stringent that, even after some relaxation, disconcertingly small numbers of cases remained in the critical groups. There were 27 operatees and 6 controls in this study. Of the 27 operatees, venous ligations were performed on 12, transorbital lobotomies on 9, thalamotomies on 2, thermocoagulation on 2, topectomy after lobotomy on 1, and enlargement of cortical ablation on 1. All of the 33 cases were schizophrenics.

Not only was the control group small, but for reasons apparently beyond the control of the study, it was poorly matched with the operated group. Advance prognostic ratings on a four-point scale were provided for 31. The average rating for the 6 control cases was 1.8 and for the 25 operatees 2.8. The control group, then, initially appeared to have a better chance for recovery. In addition, the IQ of the operated group was 80.8 with a range of 55 to 114, whereas the IQ of the control group was 100.3 with a range of 83 to 115. The incidence of recovery was small for the entire group; only two of the operatees and one of the controls was considered well enough to warrant parole.

Extensive batteries of psychometric tests were administered at four different times, two preoperative, approximately 60 days and 30 days before the operation, and two postoperative, ranging from ten to 180 days after. The tests included the Wechsler-Bellevue, the Porteus Maze, the Weigl sorting test and tests of incidental memory, verbal directions, shifting set, vestibular function, and olfaction, among many others. During the period covered by these examinations little or no amelioration in mental illness was evident. Performances on the psychometric examinations in this study bear no identifiable relationship to ultimate clinical status.

In spite of these and other problems working against the effort to reduce the ambiguities of the results, the study has considerable value. To begin with, the activities and findings

of all the disciplines are carefully documented. Many new tests were tried and some of them provide promising leads for further development. Even with the small number of cases involved, the evidence is quite convincing that the various psychosurgical insults immediately impair many psychometric and psychophysiological performances. However, with very few exceptions, the manifestations of impairment disappear with time. Among the exceptions to this trend were the inability to benefit from retesting on the Wechsler-Bellevue performance scale and on the Porteus Maze test.

These exceptions are so few in comparison to the number of tests used as to lead the authors to generalize that *all* the psychological test performance changes resulting from psychosurgery are nonspecific and transitory (p. 315). While such a conclusion stretches by only a little the description of the findings of the study, the magnitude of the testing effort may encourage many to believe that the psychosurgical procedures are without some irretrievable cost to the personality organization and performance capabilities of the patient. It is interesting to point out that this question was neither asked by the investigators nor was a suitable experimental design incorporated into the study to provide a definitive answer. While the authors usually take the conservative position that their study merely provides no evidence for such a cost, they are confronted with the paradox that such losses have been reported elsewhere for patients with frontal lobe lesions uncomplicated by psychosis. Note the following quotation:

The finding of no permanent loss in intelligence resulting from psychosurgery of these varieties has been confirmed.

Indeed the puzzling question at the present instant is why nonspecific injury, such as gunshot wounds, pathology of the frontal lobes (neoplasms) or translateral frontal lobotomy (involving the agranular cortex) is frequently attended by a permanent loss of some part of intellectual efficiency as evidenced by intelligence tests. No simple or apparent explanation of this discrepancy is available (p. 314).

The reviewer believes that the paradox is explicable in terms to which the investigators make only passing reference. They recognize that the preoperative test performances of the operatees and the initial test performances of the control group were handicapped by psychosis. It will be recalled that the average IQ of the operated group before psychosurgery was 80.8, the range 55 to 114. The scores of the operatees before psychosurgery appear to be unduly depressed. The authors, in another context, take cognizance of such a possibility in the statement:

... it is apparent that the changes due to recovery from psychosis are mainly if not entirely in the realm of affective attitudes, loss of anxiety, decrease in complaints and the like. *These affective attitudes probably interfere with the quality and quantity of performance in most of the psychologic tests that have been employed in both this and the previous study* ... (p. 278, italics mine).

They then justify ignoring this important fact in the evaluation of their findings by continuing with the statement:

... but that interference has affected the operated and control groups alike (p. 278).

To be sure, the interference *initially* may have affected the operated and control groups alike, but did they affect them equally *after* the operation? We may suspect that for many pa-

tients the surgical procedures minimized the influence of "interfering affective attitudes" of the psychosis in greater or lesser degree. This effect could well obscure performance impairment expected from brain damage and reported in cases uncomplicated by functional psychoses. For these reasons, the finding that the varieties of psychosurgery used in this study resulted in "no permanent loss in intelligence" may be true only for a schizophrenic population, and at that for one with the age and institutionalization characteristics closely resembling the one studied.

Of course, the foregoing considerations are speculative, but in view of the conflict between these Greystone findings and those contained in a substantial literature on frontal lobe damage it would be premature to conclude from the present study that the psychosurgical procedures utilized will not permanently impair psychological performances. The reviewer fears that the very considerable amount of effort which has been invested in this study will lend weight to the premature conclusion and encourage the extension of psychosurgical procedures to a wide variety of psychiatric categories on the unestablished premise that they can do no permanent harm.

JOSEPH E. BARMACK.
College of the City of New York.

WOLFF, CHARLOTTE. *The hand in psychological diagnosis.* New York: Philosophical Library, 1952. Pp. xv+218. $7.50.

The clinician who approaches this work with the anticipation that he will find a careful, objective analysis of diagnostic indicators and a scientific evaluation of their predictive limits will be disappointed. Nor will he find its theoretical treatment particularly stimulating since it is based on a vague and outmoded "faculty psychology. A brief quotation ma illustrate: "This use of the for finger transmits the human facul of discrimination, which is an esse tial quality of intelligence." The boc unfortunately is "clinical" in the ou dated, derogatory sense of the ter

WILLIAM A. HUNT.
Northwestern University.

BERGLER, EDMUND. *The supereg* New York: Grune & Stratto 1952. Pp. x+367. $6.75.

Writing a review of a psychoana lytic book that purports to be scien tific involves pronounced occupa tional hazards, particularly when th author has posted clear warning sig concerning the unconscious motiva tions of writers, reviewers, and scien tists. However—

In *The Superego*, Bergler summar izes and sharpens his psychoanalytic thinking of the last twenty year and presents numerous wide-ranging and depth-centered theoretical con structs. He valiantly attempts to go even deeper than the usual psy choanalytic probing, and to uncover not merely human defenses and resis tances, but what he calls the "de fense against a defense."

Beginning with the assumptions that "restrictions emanating from the superego . . . are biological" and that "psychic masochism is the uni versal human 'trait,'" Dr. Bergler insists that "every neurotic symptom has a five-layer structure, only the unconscious defense against the de fense becoming visible. The sequence is: (1) unconscious wish emerges; (2) superego objects (Veto #1); (3) first defense is presented by uncon scious ego; (4) superego objects once more (Veto #2); (5) second defense is presented by the unconscious ego, guilt is accepted for the lesser crime, reverberations reach the psychic sur-

face, cloaked in rationalizations. The conflict is unconscious, of infantile origin and repetitive. The end result of the infantile conflict is always oral-masochistic. Every neurosis represents a rescue-attempt from the oral danger" (p. 64).

This five-layer structure is then used by Bergler to explain such human behavior as creative writing, an cynicism, hypocrisy, pessimism, optimism, pessimo-optimism, wit, self-pity, silence, talkativism, inhibition, exhibitionism, raconteurism, thisism, thatism, and, at least by implication, kitchen sinkism. That there might possibly be some *conscious* or *one-* or *two*-layer unconscious reasons why people should be or do anything whatever, Bergler does not admit.

Never a man to mince words or qualify a dogmatism, Bergler not only outjuggles the Freudians at their deepest depths, but frequently supplies his own italics: "The end result of the infantile conflict in *every* neurosis is exclusively stabilization on the *oral* rejection level which is the basis of psychic masochism" (p. 47). "*Every* joke, in short, has as its butt the inner conscience" (p. 138). "The *severity of the super-ego*—the inner torture machine—is *unchangeable* because no method has been, or can be, invented to shorten the maturation time of the human child" (p. 325).

What evidence, it may be asked, does Bergler present to substantiate his numberless hypotheses? As far as could be determined from intensive reading of The Superego, no evidence but: (a) a handful of fragmentary case illustrations with interpretations clearly slanted by the author; (b) several references to novels, plays, poems, and biographies, with equally controversial interpretations by Bergler; (c) some charming but dubiously relevant

anecdotes; and (d) several direct and implicational appeals to the authority of Sigmund Freud. About Bergler's hypotheses two more alternate hypotheses may be offered: (a) that they constitute brilliant insights into the innermost reaches of the human personality; or (b) that they do not. If there is any scientific evidence to support the first of these alternatives it is not presented in The Superego.

ALBERT ELLIS.

New York City.

VICTOR, F. *Handwriting.* Springfield, Ill.: Charles C Thomas, 1952. Pp. xi+149. $4.00.

REVIEWED BY
WERNER WOLFF
Bard College

The fact that since 1947 a major study on handwriting has been published each year by a psychologist indicates that American psychologists become more and more interested in handwriting as a diagnostic tool. Although graphodiagnostics in Europe overshadows any other kind of psychodiagnostics, such as the Rorschach, paper-and-pencil testing, or questionnaire inquiry, graphology found great resistance in the United States. One of the reasons was that it had lent itself too much to a mixture of basic graphological observations with speculative theories on personality.

For the first time Victor makes the important point that basic psychological observation, which he calls "basic graphology," is a science of objective observation and can be independent of all graphological theories which, though interesting for any theory of personality, belong to what he calls "applied psychology." It is the opinion of the present reviewer that the author would have better used the term "applied graph-

ology" instead of "basic graphology," and "theoretical graphology" instead of "applied graphology."

Victor calls handwriting a personality projection, considering the letters as somewhat similar to an inkblot. Although, as the present reviewer observed, letters may become a screen for even a projection of images, handwriting seems to him an expression rather than a projection. As a direct expression of inner motions in outer motions, handwriting is a graphogram, more similar to an electroencephalogram than to a Rorschach inkblot.

Handwriting is the recording of inner psychological movements, of which Victor considers as the basic ones release, tension and output of energy. He compares in a suggestive way these different forms of rhythm with rhythmical forms in poetry.

On one hand, the author makes the point that handwriting as a dynamic movement does not allow any fixed correlation between graphic forms and personality traits. On the other hand, he tries to establish a correlation between handwriting and constitutional types according to Sheldon. The author should have elaborated upon this double aspect. Handwriting is both form and expression, frozen gesture and flowing motion, and it is from the relationship of both that the personal equation results.

Victor's book presents a good condensation of graphological concepts. However, it does not present any new observations and does not emphasize enough the need for graphological experiments and for a systematic correlation between elements of graphic expression and definable temperamental, clinical, and professional entities. The author's justified call for a basic graphology can be solved only if basic graphological elements can be correlated with basic mental and emotional patterns to be formulated by a scoring system. This, ultimately, will make graphology the most important psychodiagnostic instrument because the diagnosis by handwriting is the only tool which does not require the client's, or patient's, presence, and the diagnosis deals with the direct personality record itself, without using any foreign material such as inkblots or picture cards. With the proper technique it will take the least effort for a major output.

REVIEWED BY
JULIAN B. ROTTER
Ohio State University

Most serious students of projective methods of personality measurement would at least concede that if factors of handedness, conditions of sampling, and special training are controlled, better than chance relationship may be found between some aspects of handwriting and similar aspects of other motor tasks. They might also agree that handwriting may reveal at a better than chance level such characteristics as the educational attainment or experience of the subject. In some cases markedly individualistic writing may allow for the "diagnosis" of some generalized and significant personality characteristics. Speaking generally, however, American psychologists have refused to accept the thesis that handwriting may be confidently used as an efficient, reliable, or valid method of predicting significant personality variables. Victor sets as his goal for this book to promote " . . . as wide, serious and efficient a use of handwriting analysis in this country as it enjoys abroad" (p. 3).

The author feels that part of the handicap of graphology has been that it has been tied too closely to various conflicting schools of psychology, and he advocates a separation of what is "basic" graphology from "applied" graphology. He sees handwriting as a kind of recorded rhythm, and stresses movement concepts as opposed to a static graphology which deals with signs, the shape of specific letters, etc. The movement of the hand on or above the paper, connectedness, and relationship of letters are of major importance. Handwriting analysis deals for the most part with these characteristics of movement, pressure, direction, and the symbolism of space. Some idea of the nature and implications of this type of analysis may be obtained from the following quotations. In the paragraphs below the author is dealing with the problem of depth rhythm, using an analogy of a water current:

Soon the stream, and with it the stroke, may become sluggish. Their borders are no longer sharp. The written letters show fringes on one or both sides (Figure 6a), as if written on blotting paper. Just as the water runs slowly through meadows and swamps, without bed or border, the life stream of the writer loses its intensity, its liveliness and "directed energy." Having lost its aim, it becomes wasted.

The river bed may dry up. Although the direction of the river's path is recognizable, the water itself, the energy, is gone (the dry gullies in the summer desert). Handwriting analogous to this simulates strength, its borders insufficiently filled with ink, or even bare of it (Figure 8i).

On its way, the river may find a dam in its path. Similarly, handwriting can be dammed up by sudden stops of the pen which retard the forward movement (tense rhythm). When the dam breaks, however, the suddenly released energy

of the river as well as of the emotions will cause untold, unforseeable damage. Heavy, suddenly-stopped strokes always carry in them the danger of a break, of eruption and devastation (p. 47).

In dealing with the problem of direction the author has the following to say:

Where the writing starts, the mind symbolically or unconsciously experiences the *beginning*, the origin; and origin stands for the Mother. The beginning becomes the past just as soon as the writer leaves it on his way forward. He moves away from the past into the future.

The left side, then, represents the past. "Mother" and "left" become interchangeable. To live, in the Occidental way of thinking, means to keep moving forward, toward action, freedom and independence: we move to the right. But at the right is found the antithesis of the mother, he who represents all these characteristics: Father. The right side, thus, becomes identified with the father. These images, of father and mother, influence everyone's life, and their effect on the writer's personality finds its expression in a greater or lesser tendency to favor the left or right (p. 52).

Applied graphology applies the principles of handwriting analysis to any of several personality theories or problems primarily on the basis of analogy. A sample case of graphological analysis is made, and this is translated into personality descriptions in terms of the constructs of Jung, Freud, Horney, and Sheldon with equal facility. In discussing applied graphology, although Victor defends the graphologist's difficulty in determining sex (which, after all shouldn't be difficult since attitudes toward father and mother should at least be related to sex) he does imply that graphology can contribute to the diagnosis of a great many problems providing the referrer states

clearly the problems which he wishes to be answered. The quotation below illustrates some of the potential range of usefulness of handwriting:

All questions must be precise; the analyst should know whether he is to search for criminal tendencies, for instance, or for neuroses, psychoses, or whatever. If a doctor wishes to know whether the subject is a "bleeder" (hemophiliac), he must at least supply to the analyst the characteristics of this particular disease, and if the doctor wishes information on the progress of a treatment he must be prepared to give all pertinent facts. Even if the graphologist cannot give a diagnosis, he will be able to point out symptoms that may be conclusive enough to the medical man (p. 92).

As to the verification of all this, the author does not believe that much is to be obtained from "mass experimentation" and even when stating that the relationship of middle, lower, and upper zones should be balanced in a 1:1:1 ratio which would be the proportion of idealism, to judgment and feeling, to instinct, he states that this ratio cannot be expressed in inches but that the analyst's eye must be trained to recognize it in the script. Since the author does not feel that it would be profitable to come to grips with the American tendency to insist upon objectivity, quantification and experimental tests, it is the reviewer's impression that this book will fail in its goal of obtaining wide acceptance of handwriting analysis in this country. It should be said in all fairness to the author, however, that the reviewer's handwriting reveals a veritable bonanza of pathologies and produces considerable aggression in his secretary. For the reviewer, the typewriter was a wonderful invention.

ABRAMSON, H. A. (Ed.) *Problems of consciousness.* Transactions of the third conference. New York: Josiah Macy, Jr. Foundation, 1952. Pp. 156. $3.25.

BAL, ALEXANDRE. *L'attention et ses maladies.* Que sais-je? #541. Paris: Presses Universitaires de France, 1952. Pp. 123.

BELLAK, LEOPOLD. *Psychology of physical illness.* New York: Grune & Stratton, 1952. Pp. vii+243. $5.50.

BINGHAM, WALTER V., D. BORING, EDWIN G., BURT, CYRIL, ELLIOTT, RICHARD M., GEMELLI, AGOSTINO, GESELL, ARNOLD, HULL, CLARK L., HUNTER, WALTER S., KATZ DAVID, MICHOTTE, ALBERT, PIAGET, JEAN, PIÉRON, HENRI, THOMSON, GODFREY, THURSTONE, L. L., TOLMAN, EDWARD C., LANGFELD, HERBERT S., BORING, EDWIN G., WERNER, HEINZ, & YERKES, ROBERT M. (Eds.) *A history of psychology in autobiography.* Vol. IV. Worcester: Clark Univer. Press. 1952. Pp. xii+356. $7.50.

BRENNAN, ROBERT E. *General psychology.* (2nd Ed.) New York: Macmillan, 1952. Pp. xxxii+524. $5.50.

BOWERS, HENRY. *Research in the training of teachers.* Toronto: Dent & Sons, 1952. Pp. vii+167. $1.90.

BROWER, DANIEL, & ABT, LAWRENCE E. (Eds.) *Progress in clinical psychology.* New York: Grune & Stratton, 1952. Pp. xi+328. $5.75.

COOMBS, CLYDE H. *A theory of psychological scaling.* Ann Arbor: Engineering Research Inst., Univer. of Mich., 1952. Pp. vi+94. $1.75.

DE GRAZIA, SEBASTIAN. *Errors of psychotherapy.* New York: Doubleday, 1952. Pp. 288. $3.00.

FEDERN, P., & WEISS, E. (Eds.) *Ego psychology and the psychoses.* New York: Basic Books, 1953. Pp. 375. $6.00.

FIEDLER, MIRIAM FORSTER. *Deaf children in a hearing world.* New York: Ronald, 1952. Pp. viii+320. $5.00.

FLANAGAN, JOHN C., SANFORD, FILLMORE H., MACMILLAN, JOHN W., KENNEDY, JOHN L., MELTON, ARTHUR W., WILLIAMS, FREDERICK W., BAIER, DONALD E., & FINCH, GLEN. *Current trends—psychology in the world emergency.* Pittsburgh: Univer. of Pittsburgh Press, 1952. Pp. 198. $4.00.

GILBERT, JEANNE G. *Understanding old age.* New York: Ronald Press, 1952. Pp. ix+422. $5.00.

GRIFFITHS, WIILLAM. *Behavior difficulties of children as perceived and judged by parents, teachers, and children themselves.* Minneapolis: Univer. of Minnesota Press, 1952. Pp. xii+116. $3.00.

HALL, B. H., GANGEMI, M., NORRIS, V. L., VAIL, V. H., SAWATSKY, G. *Psychiatric aide education.* New York: Grune & Stratton, 1952. Pp. xvi+168. $5.75.

HARTLEY, EUGENE L., & HARTLEY, RUTH E. *Fundamentals of social psychology.* New York: Knopf, 1952. Pp. xix+740. $5.50.

HAYEK, F. A. *The sensory order.* Chicago: Univer. of Chicago Press, 1952. Pp. xxii+209. $5.50.

HILGARD, E. R., KUBIE, L. S., & PUMPIAN-MINDLIN, E. (Eds.) *Psychoanalysis as science.* The Hixon Lectures on the Scientific Status of Psychoanalysis delivered at the California Institute of Technology. Stanford: Stanford Univer. Press, 1952. Pp. x+174. $4.25.

HOOKER, DAVENPORT. *The prenatal origin of behavior.* Lawrence:

Univer. of Kansas Press, 1952. Pp. viii+143. $2.50.

HOYLES. J. A. *The treatment of the young delinquent.* New York: Philosophical Library, 1952. Pp. vii+273. $4.75.

HULL, CLARK L. *A behavior system.* New Haven: Yale Univer. Press, 1952. Pp. ix+372. $6.00.

ITTELSON, WILLIAM H. *The Ames demonstrations in perception.* Princeton: Princeton Univer. Press, 1952. Pp. xvi+88. $4.00.

JOHNSON, W., DARLEY, F. L., & SPRIESTERSBACH, D. C. *Diagnostic manual in speech correction.* New York: Harper, 1952. Pp. viii+221. $2.50.

KLEIN, MELANIE, HEIMANN, PAULD, ISAACS, SUSAN, & RIVIERA, JOAN. (Eds.) *Developments in psychoanalysis.* London: Hogarth, 1952. Pp. viii+368. 30s.

KOCH, CHARLES. *The tree test.* New York: Grune & Stratton, 1952. Pp. 87. $4.50.

KOHLER, FRED. *Evolution and human destiny.* New York: Philosophical Library, 1952. Pp. 120. $2.75.

NORTHWAY, MARY L. *A primer of sociometry.* Toronto: Univer. of Toronto Press, 1952. Pp. vi+48. $2.25.

REIK, THEODOR. *The secret self.* New York: Farrar, Straus & Young, 1952. Pp. 329. $3.50.

RIESEN, AUSTIN H., & KINDER, ELAINE F. *The postural development of infant chimpanzees.* New Haven: Yale Univer. Press, 1952. Pp. xx+204. $5.00.

SHAFFER, G. WILSON, & LAZARUS, RICHARD S. *Fundamental concepts in clinical psychology.* New York: McGraw-Hill, 1952. Pp. xi+540. $6.00.

SZONDI, L. *Experimental diagnostics of drives.* New York: Grune & Stratton, 1952. Pp. x+220. $13.50.

TANSLEY, ARTHUR G. *Mind and life.* New York: John DeGraff, 1952. Pp. ix+171. $3.50.

TRENAMAN, J. *Out of step.* New York: Philosophical Library, 1952. Pp. xx+223. $4.75.

VERNON, M. D. *A further study of visual perception.* New York: Cambridge Univer. Press, 1952. Pp. xi+289. $7.00.

WALLON, H. & ÉVART-CHMIELNISKI, E. *Les mécanismes de la mémoire en rapport avec ses objets.* Paris: Presses Universitaires de France, 1951. Pp. viii+116. 400fr.

WISDOM, JOHN. *Other minds.* New York: Philosophical Library, 1952. Pp. 259. $4.75.

WOLFLE, DAEL, BUXTON, CLAUDE E., COFER, CHARLES N., GUSTAD, JOHN W., MacLEOD, ROBERT B., & McKEACHIE, WLBERT J. *Improving undergraduate instruction in psychology.* New York: Macmillan, 1952. Pp. vii+60. $1.25.

Psychological Bulletin

A HISTORY OF INTROSPECTION

EDWIN G. BORING

Harvard University

A proper but cumbersome title for this article would be "The History of the Availability of Consciousness to Observation in Scientific Psychology." If conscious experience can be said to exist, then the question arises as to whether modern psychology ought not to take into consideration its data, as indeed it used always to do. Thus my paper might even be called "What Became of Introspection?" One common answer to that question would be that introspection was not viable and so gradually became extinct. Another answer, however, is that introspection is still with us, doing its business under various aliases, of which *verbal report* is one. The former statement about the failure of introspection is approximately true of that introspection which flourished under Titchener at Cornell in 1900–1920, whereas the latter statement about camouflaged introspection is accepted by the modern positivists who hold that the concept of conscious experience has meaning only when it is defined operationally.

DUALISM

The belief in the existence of conscious mind in man is very old, as old as philosophy and as old as the belief in the immortality of the soul, the immortality of that part of a person that is not his mortal body. Thus it has come about that something conscious is usually one term in a dual-ism, like mind *vs.* matter, the rational *vs.* the irrational, or purpose *vs.* mechanism. There have been psychological monists, like La Mettrie (44), the materialist, who argued in 1748 that man is a machine and who got himself consequently into theological trouble, but even he was more concerned with reducing to their bodily bases the mental states that dualism had already established than in describing man without benefit of dualism.

Inevitably the doctrine of immortality and the old-time importance of theology played a role in psychology. The words for soul and mind are not distinguished in French and German (*l'âme, Seele*) nor are the Greek and Latin words (*psyche, nous; anima, mens*) as distinct as the English translations. It was the faculty of reason that carried with it the right to immortality, and Descartes, a devout Catholic, gave men rational souls, made of unextended immortal substance, and maintained that animals are mortal irrational automata (20). Thus Descartes became an important ancestor in both the dualistic (conscious, introspective) line of descent, and in the objective (mechanistic, reflex, tropistic) line.

British empiricism fixed dualism and the concept of consciousness upon psychology. Locke, Berkeley, Hume, Hartley, Reid, Stewart, Thomas Brown, the two Mills, and

Bain, all were concerned in different ways with how the mind gets to know about the external world. Thus they recognized the basic mind-matter dichotomy. Presently there came also into the hands of these philosophers the doctrine of association which dealt with the synthetic relations among the items of mind or consciousness (8, pp. 157–245). There never was—nor is there now—a good word for this immaterial term of the mind-matter dichotomy. James was complaining about that in 1890 (32, I, pp. 185–187). Mostly the word was either *mind* (*Seele*) or *consciousness* (*Bewusstheit*). Nineteenth-century psychology formulated the dichotomy as psychophysical parallelism, and that doctrine was so firmly impressed upon psychological thinking that the American operational revolution of the present century came about only with the greatest difficulty.

It would not be profitable to go into great detail here about the history of the belief in what we are calling *consciousness*. The existence of consciousness seemed for many centuries to be an obvious immediate datum, the basic undeniable reality of one's own existence. "Cogito, ergo sum," said Descartes. James summed the matter up (32, I, p. 185):

Introspective Observation is what we have to rely on first and foremost and always. The word introspection needs hardly to be defined—it means, of course, looking into our own minds and reporting what we there discover. *Every one agrees that we there discover states of consciousness.* So far as I know, the existence of such states has never been doubted by any critic, however skeptical in other respects he may have been. That we have *cogitations* of some sort is the *inconcussum* in a world most of whose other facts have at some time tottered in the breath of philosophical doubt. All people unhesitatingly believe that they feel

themselves thinking, and that they distinguish the mental state as an inward activity or passion, from all the objects with which it may cognitively deal. I regard this belief as the most fundamental of all the postulates of Psychology, and shall discard all curious inquiries about its certainty as too metaphysical for the scope of this book.

In general the philosophers, physiologists, and physicists who founded the new experimental psychology in 1850–1870—Fechner, Lotze, Helmholtz, Wundt, Hering, Mach, and their associates—were psychophysical parallelists who would have subscribed to this view of James' (8, pp. 261–356). Psychology—even the new "physiological psychology"—was essentially the study of consciousness, and its chief method was introspection. Physiology came in because these parallelists believed in "no psychosis without neurosis" (Huxley's phrase, 30, 1874) and thus could employ the apparatus of the physiological laboratory to control stimuli and to record the effects of neural events.

About introspection (*innere Wahrnehmung*) there was, however, some question. There is a long history of opinions on the manner in which the mind observes its own processes, one that begins with Aristotle and Plato and carries on to the present. Eisler has abstracted the views of eighty-four writers on the subject, from Aristotle to the beginning of the present century (21, III, pp. 1735–1742). Locke, founding empiricism, held that all ideas—that is to say, the contents of the mind—come from experience either by sensation, which provides knowledge of the external world, or by reflection, which is the inner sense and provides knowledge of the mind's own operations. Neither sensation nor reflection, however, was regarded by the early empiricist

as a process subject to error. The belief grew up that to have conscious experience is also to know that you have it, and thus ultimately Wundt, basing his new systematic physiological psychology upon British empiricism, defined introspection as immediate experience (**98**, pp. 1–6). The facts of physical science, he thought, are mediated and derived by inference from immediate experience, which in and of itself is immediately given and constitutes the subject matter of psychology. This view suggests that Wundt thought that introspection cannot lie, but actually there was an inconsistency there, for the Wundtian laboratory put great emphasis upon training in introspective observation and in the accurate description of consciousness.

Brentano wrote in 1874: "The phenomena inwardly apprehended are true in themselves. As they appear ... so they are in reality. Who then can deny that in this a great superiority of psychology over the physical sciences comes to light?" (**12**, I, pp. 131–203). Against this view, James remarked: "If to *have* feelings or thoughts in their immediacy were enough, babies in the cradle would be psychologists, and infallible ones" (**32**, I, p. 189). The classical objection to the *ipso facto* adequacy of the immediate was raised by Auguste Comte, the founder of positivism, who noted that introspection, being an activity of the mind, would always find the mind introspecting and never engaged in the great variety of its other activities (**17**, p. 64). Actually Comte's argument was, however, much more than this quibble, which could have been answered by the statement that introspection is not a procedure but merely the recognition that knowledge, when given, exists as knowledge. Comte was complaining, as did twentieth-century behaviorists, that introspection is unreliable, that it results in descriptions which often cannot be verified, and that in many other ways it fails of the positive character that science demands.

J. S. Mill answered Comte's quibble by asserting that introspection is a process and requires training for reliability. It is not strictly immediate, Mill thought, for it involves memory —immediate memory, perhaps; yet immediate memory is not the datum itself and comes with a chance for error in it (**53**, p. 64). On this whole matter, see James' excellent discussion (**32**, I, pp. 187–192). Mill's point is reinforced by the modern realization that it is almost impossible to distinguish between anesthesia and immediate anterograde amnesia: a man whose memory lasts only one second is so crippled in capacity for introspection as to be practically as unconscious as any reacting organism or machine.

CLASSICAL INTROSPECTION

We may regard that introspection as classical which was defined by fairly formal rules and principles and which directly emerged from the early practices in Wundt's laboratory at Leipzig. Of course, there were no immutable rules for introspection. The great men kept disagreeing with one another and changing their minds. Nevertheless there was a body of opinion which was in general shared by Wundt, by Külpe before he left Leipzig, by G. E. Müller at Göttingen, by Titchener at Cornell and by many other less important "introspectionists" who accepted the leadership of these men. Stumpf at Berlin held to less constrained principles, and Külpe's later doctrine of introspection after he had gone to Würzburg was opposed by Wundt and Titchener. Classical introspection is the com-

mon belief that the description of consciousness reveals complexes that are constituted of patterns of sensory elements. It was against this doctrine that Külpe at Würzburg, the behaviorists under Watson and the Gestalt psychologists at Wertheimer's initiative revolted. Introspection got its *ism* because these protesting new schools needed a clear and stable contrasting background against which to exhibit their novel features. No proponent of introspection as the basic method of psychology ever called himself an *introspectionist*. Usually he called himself a *psychologist*.

Wundt, undertaking to establish the new psychology as a science, turned to chemistry for his model. This choice landed him in elementism, with associationism to provide for synthesis. The psychological atoms were thus sensations and perhaps also feelings and images. The psychological molecules were precepptions and ideas (*Vorstellungen*) and the more complex combinations (*Verbindungen*). Because Wundt changed his views from time to time about images and feelings, the sensation became the example of the sort of stuff that appears in a good description of consciousness. Thus, half a century later, we find Titchener concluding that *sensory* is the adjective that best indicates the nature of the contents of consciousness (**85**, pp. 259–268). In this way Wundt fixed both elementism and sensationism upon introspection, and introspectionism in the proper laboratories always yielded sensory elements because that was "good" observation. It seems reasonable to suppose that laboratory atmosphere and local cultural tradition did more to perpetuate this value than did any published admonitions about observation.

Although Wundt defined the sub-

ject matter of psychology as immediate experience (**97**; **98**, pp. 1–6), he did distinguish introspection (*Selbstbeobachtung*) from inner perception (*innere Wahrnehmung*). Inner perception might be self-validating, but it was not science. Wundt insisted on the training of observers. Even in the reaction experiment Leipzig observers had to be trained to perform the prescribed acts in perception, apperception, cognition, discrimination, judgment, choice, and the like, and to report when consciousness deviated from what had been called for. Thus it is said that no observer who had performed less than 10,000 of these introspectively controlled reactions was suitable to provide data for published research from Wundt's laboratory. Some Americans, like Cattell, had the idea that the minds of untrained observers might also be of interest to psychology, and later a bitter little quarrel on this matter developed between Baldwin and Titchener (**8**, pp. 413 f., 555). For all that, Wundt's notion of what constitutes proper introspection was much more liberal than is generally supposed, for he left room in formal introspection for retrospection and for indirect report. He was much less flexible in respect of the elements and their sensory nature.

What happened next to introspection was the acceptance of the conception that physics and psychology differ from each other in points of view but not in fundamental materials. Mach in 1886 argued that experience ("sensation") is the subject matter of all the sciences (**48**), and Avenarius a few years later that psychology views experience as dependent upon the functioning of the nervous system (he called it the "System C") and physics as independent of the action of the nervous

system (3). Presently, after the two men had agreed that they agreed, they had great influence upon Külpe and Titchener who were both then at Leipzig. In his textbook of 1893 Külpe accepted this distinction by point of view (41, pp. 9–13), but Titchener is the person who emphasized it most. In 1910, he was saying that the data of introspection are "the sum-total of human experience considered as dependent upon the experiencing person" (79, pp. 1–25), and later he could write the formula:

Introspection = psychological
(clear experience →report),

which means that introspection is the having of clear experience under the psychological point of view and the reporting upon it also under the psychological point of view (83, pp. 1–26). Substitute physical for psychological, and you have the formula for physics. The stock example for introspection is the illusion, the case where perception differs from stimulus-object in some respect. For perception experience is regarded just as it comes, dependent upon the perceiving of the perceiving person and thus the action of his nervous system. For the physical account of the object, however, the perceiver must be abstracted from and the physicist has resort to measurement and other physical technics. Titchener held to this distinction by point of view all his life (85, pp. 259–268).

It was Külpe who split Wundt's psychological atom, analyzing sensation into its four inseparable but independently variable attributes: quality, intensity, extensity, and duration (41, pp. 30–38). Titchener later held to this view which served to tighten rather than to loosen the constraints of atomism upon introspective psychology (6, pp. 17–35).

One of the most thorough discussions of introspection was provided by the erudite G. E. Müller in 1911 (55, pp. 61–176). Müller was more liberal than Wundt and left room for all the indirect and retrospective forms of introspection. Being primarily interested in the application of introspection to memory, he distinguished, for instance, between the present recall of the past apperception of a past event and the present apperception of the present recall of a past event, an important distinction, since present apperception can be interrogated as to detail whereas past apperception has become fixed and no longer subject to exploration.

It was Titchener who placed the greatest constraints upon introspection by his requirement that the description of consciousness should exclude statements of meaning. At first Titchener had perception in mind and called the report of meanings the *stimulus-error*, insisting that trained observers by taking the psychological point of view would describe consciousness ("dependent experience") and attempt no statements about the stimulus-objects ("independent experience" as given by the point of view of physics) (5; 79, pp. 202 f.). After Külpe had claimed to find imageless (non-sensory) thoughts in the consciousnesses of judgment, action, and other thought processes, Titchener broadened his criticism to an objection against the inclusion of any meanings at all in the data of introspection (80). He was arguing that straight description (*Beschreibung, cognitio rei*) would yield the kind of sensory contents that had become standard in classical introspection, and that inferences about conscious data (*Kundgabe, cognitio circa rem*) are meanings which do not exist as do the observed sensory processes (81,

82).   Thus his psychology has even been called *existential psychology*, because he believed that the meanings, occurring as inferences, lack the positive character of sensations and images, the existential data (**85**, p. 138).

It was never wholly true that introspection was photographic and not elaborated by inferences or meanings. Reference to typical introspective researches from Titchener's laboratory establishes this point (**28, 58, 25, 64, 59, 16, 31**). There was too much dependence upon retrospection. It could take twenty minutes to describe the conscious content of a second and a half and at the end of that period the observer was cudgeling his brain to recall what had actually happened more than a thousand seconds ago, relying, of course, on inference. At the Yale meeting of the APA in 1913, J. W. Baird with great enthusiasm arranged for a public demonstration of introspection with the trained observers from his laboratory at Clark, but the performance was not impressive. Introspection with inference and meaning left out as much as possible becomes a dull taxonomic account of sensory events which, since they suggest almost no functional value for the organism, are peculiarly uninteresting to the American scientific temper.

Classical introspection, it seems to me, went out of style after Titchener's death (1927) because it had demonstrated no functional use and therefore seemed dull, and also because it was unreliable. Laboratory atmosphere crept into the descriptions, and it was not possible to verify, from one laboratory to another, the introspective accounts of the consciousnesses of action, feeling, choice, and judgment. It is not surprising, therefore, that Külpe, Watson and Wertheimer, all within a decade (1904–1913), reacted vigorously against the constraints of this idealistic but rigid pedantry.

DESCRIPTION OF THE IMPALPABLE

What came to be called *systematic experimental introspection* developed at Würzburg in 1901–1905 under Külpe's leadership (**8**, pp. 401–410, 433–435). Külpe, influenced like Titchener toward positivism by Mach, had gone from Leipzig to Würzburg with the conviction that experimental psychology ought to do something about thought. The new experimental psychology could handle sensation, perception and reaction, and Ebbinghaus in 1885 had added memory to its repertoire. Wundt had said that thought could not be studied experimentally, but Külpe, positivist, was convinced that all you had to do was to get observers thinking under controlled conditions and then have them introspect upon the thought process.

There followed a brilliant series of papers by Külpe's students: Mayer and Orth on association (1901), Marbe on judgment (1901), Orth on feeling (1903), Watt on thought (1905), Ach on action and thought (1905). Every one of these investigators found what we have called classical introspection inadequate to his problem. Mayer and Orth could describe the associated trains of images that run on in thinking but could discover from introspection no clue as to how thought is directed toward a goal (**50**). Marbe found judgments forming readily in terms of images, but got from introspection no hint as to how or why they were formed (**49**). Feeling resisted Orth's introspective analysis and he was obliged to invent a vague term, *conscious attitude*, to describe the affective life. Certainly feelings did not

appear as sensations or images to his observers (60). Watt and Ach worked independently and came to mutually consistent conclusions. Watt, to make introspection more efficient, invented fractionation. He split up the psychological event under investigation into several successive periods and investigated each by itself, thus reducing the amount of memory and inference that were involved in the introspective report. Still the essential in thought eluded him, until he realized that the goal-directedness of thinking is predetermined by the task or instruction—the *Aufgabe* he called it—which the observer accepted before the individual thought process got under way (92). Ach developed the concept of the *determining tendency* as the unconscious guide which steers the conscious processes along a predetermined course to solve whatever problem thought is directed upon. He also elaborated fractionation with chronoscope control and coined the phrase *systematic experimental introspection*. The determining tendency itself is unconscious, but the conscious processes which it directs seemed to Ach's observers not to be describable in the terms of classical introspection, that is to say, in images and sensations. Ach therefore invented the term *awareness* for these vague and elusive contents of consciousness and his observers learned to describe their consciousnesses in terms of impalpable awarenesses (*unanschauliche Bewusstheiten*) (1).

The Würzburgers thought they had discovered by introspection a new kind of mental element, but the *Bewusstheit* never gained the accepted status of a sensation or an image. Instead the Würzburgers were said to have discovered imageless thought, and many persons argued that the

school had failed because its finding was negative: thoughts were not images, but what actually were they? Titchener, however, believed he knew. He said that these Würzburg thoughts were in part conscious attitudes which are vague evanescent patterns of sensations and images, and in part meanings and inferences which ought to be kept out of psychology as the *Kundgabe* which is not true description (80). We, with the perspective of forty years upon us, see that the main contribution lay in the realization of the importance of the unconscious *Aufgabe* and determining tendency. The course of thought is unconsciously determined: that is a conclusion which fitted the *Zeitgeist* of the period of its discovery, when Freud too was discovering that motivation is ordinarily not available to introspection.

Külpe's conclusion was, however, different. He believed that the impalpable awarenesses had been established as valid data of consciousness and he called them *functions* to distinguish them from the sensations and images of classical introspection, which he called contents (43). *Funktionen* and *Inhalte* are two kinds of conscious data that make up what has been termed the bipartite psychology of Külpe's later days. In this choice Külpe was combining the introspection of Wundt with the introspection of Brentano. He was also making easier the coming protest of Gestalt psychology against Wundtian introspection.

AWARENESS OF MENTAL ACTIVITY

Meanwhile nearly all the philosophers and psychologists were dualists and most of the psychologists were also psychophysical parallelists. If you believe in conscious events as dependent upon brain events but wholly separate and different from

the brain events, then you must believe in some kind of introspection or inner perception whereby you obtain your evidence about the mental events. The behavioristic monism of the twentieth century was unknown in the nineteenth. A belief in some kind of introspection was general in psychology and also in common sense.

The appeal to introspection was especially important in the case of act psychology, which claimed that a careful and unbiased examination of the mind shows that it does not consist of stable contents like images and sensations, but of acts directed intentionally upon an object or of activities striving purposively toward a goal (8, pp. 439–456, 715–721). We have already seen that Brentano defended introspection as self-validating. He was the representative of intentionalistic act psychology who was contemporary with Wundt, and who thus posed the dilemma between Wundt's contents and his own acts (12), a dilemma of which Külpe, as we have just noted, seized both horns. Brentano influenced the philosopher James Ward in his subject-object conative psychology of 1886, revised in 1918 (87), and Ward influenced McDougall, who, in spite of having once defined psychology as the science of behavior, elaborated a purposive psychology in 1923, a system that made purpose and striving a characteristic of all mental activity (51).

In Germany, Stumpf, stimulated by Brentano's sponsorship of psychic acts and by Husserl's argument for phenomenology as the simplest description of experience (29), came to the conclusion that Wundt's kind of introspection yields the data of phenomenology but that psychology proper consists rather of Brentano's acts or, as Stumpf called them, *psychic functions* (76). Thus it is cor-

rect to say that by 1915 both Stumpf and Külpe believed in two kinds of introspective data: on the one hand, Stumpf in phenomena and Külpe in contents, and, on the other, both of them in functions (acts). Külpe was inclined to think that the functions were observed retrospectively (*rückschauende Selbstbeobachtung*), the contents immediately (*anschauende Selbstbeobachtung*) (43, pp. 42–45).

Except for Titchener and his satellites, American psychology tended all along to be practical and functional, in the Darwinian sense. As such it was destined to become behavioristic. It is interesting, therefore, to note that early American functional psychology of James, Dewey, Angell, and the Chicago school was introspective. Organisms have acquired consciousness because of its adaptive function, the argument ran. When the smooth course of habitual action is interrupted by external events, then "in steps consciousness," said James Angell, to solve the organism's problem (2; 9, pp. 276–278). It is because functional psychology regarded the data of consciousness as essential to an understanding of the adjustment of man to his environment that Watson, founding behaviorism, declared that he was as much against functional psychology as against introspectionism.

PHENOMENOLOGICAL DESCRIPTION

The next protest against the constraints of classical introspection came in connection with the founding of Gestalt psychology—by Wertheimer, we generally say, in his paper of 1912 on seen movement (94). Wertheimer was working on the conditions of visually perceived movement. You can see movement when no stimulus object moves, as when stimulus displacement is discrete. Seen movement is thus a con-

scious, not a physical, event. Classical introspection would have required the description of perceived movement with reference to conscious contents, or mental processes, or images and sensations, or perhaps the attributes of sensation. Wertheimer thought, however, that any such reference or analysis would be a supererogation. Perceived movement can be recognized as itself and its conditions studied; why bother then with the Leipzig hocus-pocus? Since seen movement can thus be accepted immediately as an identifiable phenomenon, Wertheimer called it Φ— the "Φ-phenomenon." In 1912 the notion of phenomenology was in the air. Husserl had used the term for the free unbiased description of experience ("being") (29) and Stumpf had picked it up (76). Thus Köhler and the other Gestalt psychologists came always to speak of the data of direct experience as *phenomena*, avoiding all the words that were associated with classical introspection. Later it was such *phenomenological observation* that became a technic to displace *introspection* (8, pp. 601–607).

This Magna Carta of phenomenology presently released a great deal of good research, most of it on problems of perceptions. In G. E. Müller's laboratory Katz's work on brightness constancy (34) had even preceded Wertheimer's, and Rubin's classical study of figure and ground (68) came soon after. There began a long series of investigations of the laws of perceived form, studies which introduced new descriptive concepts for the phenomena, like *organization* and *articulation*, and new functional concepts, like *closure*, *transposition*, and *object constancy* (8, pp. 611–614).

Nearly all these perceptual studies have been performed in an atmosphere of dualism. You try to find

the stimulus conditions or else the brain pattern that is necessary and sufficient for the perception. Wertheimer, Köhler, and Koffka have all supported the concept of *isomorphism*, the hypothesis that the field pattern of the perception corresponds topologically to the field pattern of the underlying events in the brain, and, while neither Gestalt psychology nor experimental phenomenology requires isomorphism as a basic concept, nevertheless isomorphism requires some kind of dualism, and thus the phenomena become one term in its psychophysiological correlation. Köhler's great book on *Physische Gestalten* in 1920 supported this view (36).

As Gestalt psychology waxed, classical introspection waned. Wertheimer's paper on phenomenal movement was in 1912 (94). Külpe died in 1915. Köhler worked with apes on the island of Teneriffe during World War I and applied the new phenomenological principles in the description of their psychology (35). Koffka's students were busy publishing papers on perception. Wundt died in 1920, the year that Köhler published *Physische Gestalten* (36). In 1922 Köhler went to Berlin to succeed Stumpf. The Gestalt psychologists had started a new journal devoted to their interests in 1921, *Psychologische Forschung*, and Wertheimer used its early pages to make the case against classical introspection (94). Koffka restated the case in English for Americans in 1922 (38). Titchener died in 1927. Köhler's *Gestalt Psychology* appeared in 1929 (37), and Koffka's *Principles* in 1935 (39). It is reasonable to say that phenomenological observation had won out over classical introspection by 1930.

Under Hitler's influence the Gestalt psychologists who remained pro-

ductive all came to America. There the victory of phenomenology, made easier by Titchener's death, was no great triumph, for other strong forces were operating to swing American psychology toward behavioristics. Nevertheless, phenomenology remained, not only respectable, but stimulating and useful in initial attacks upon many psychological problems, as Gibson's recent phenomenological study of the visual world shows (26). So here we come to a case where introspection, under an alias, can be said to be still practiced, provided the word *introspection* is not restricted to its Leipzig-Cornell meaning.

PATIENTS' PROTOCOLS

The emphasis which modern psychopathology places on the unconscious creates for it a complementary concern with the conscious. Thus psychoanalysis stresses the importance to therapy of bringing repressed ideas from the unconscious into consciousness. The analysand, bubbling free associations on the couch, is certainly giving the analyst information about his consciousness (*Kundgabe*) though he remain far from the use of classical introspection. When and how, we may ask, did psychopathology get itself concerned with the content of consciousness?

Nearly always the first evidence of what we now call mental disease lies in abnormal conduct, in maladaptive behavior. The abnormal person, witch or patient as the case may be, first calls attention to himself by queer or alarming conduct. The obvious symptoms that require social action, remedial or protective, are usually not reports of visions or complaints about voices, but such deviations from standard behavior as

inconvenience others. Nevertheless psychopathology, which grew up surrounded by a belief in dualism, was never primarily behavioristic. There was for it always the presumption that a witch is conscious, even though the devil might have taken possession of her will, and later that the hallucinations and delusions of the hysterical patient are conscious phenomena. Subjectivism, always implicit in these symptoms, was not very often explicit before the end of the nineteenth century.

Zilboorg's account makes it clear how the idea of mental derangement began in the conception of demoniacal possession (96, 99). For these possessed people and for the fools, except in those cases where they were honored, the therapy consisted of discipline, threats, fetters, and blows, none of which actually had much value except to relieve those who administered the punishment. Even the Renaissance, which is said to have "discovered man," did not free these unhappy victims of an intolerant theological self-assurance, until at last the reaction toward humane treatment arrived with Pinel and his successors early in the nineteenth century. During the seventeenth and eighteenth centuries you get as subjective data the reports of melancholy (sometimes ending in suicide), of passions, of deliriums ("errors of reason"), of fantasies, of cholers, humors and madness, of spleen, vapors and hysterical distempers, of love as a cause of mental disability. An incubus might be a woman's hallucination, delusion, or wish projection, or else a fiction of other people's belief about her. The reforms of the nineteenth century toward the humane treatment of the insane and the rise of the concept of mental disease (Pinel, 1801) did not

go far toward the subjectivization of psychopathology (**61**). Braid's theory of hypnosis, as the scientific successor to mesmerism was called, was based on suggestion as a principle, a mentalistic but not a conscious entity (**11**). Liébeault cured a patient of sciatic pain by hypnosis; is a patient who says he feels pain introspecting? Liébeault was a dualist, for the title of his book asserts that he was studying *l'action de la morale sur le physique:* a treatise on psychosomatic medicine in 1866 (**45**). Later Charcot worked out the stigmata of hysteria and thus, as he thought, of hypnosis, but most of the stigmata were not described in conscious terms, being phenomena like anesthesias, amnesias, and catatonias (**15**, III & IX). Kraepelin, Wundt's one-time student, whose classical system of mental diseases reached maturity about 1896, established the basic dichotomy between manic-depressive psychoses and dementia praecox (**40**). Thus he recognized elation, depression, and hallucinations as symptoms of mental disease, but that is a far cry from saying that his psychiatry was based on some kind of introspection.

Nevertheless this last decade of the nineteenth century was the decade for psychopathology to turn truly psychological. It marked the emergence of Janet first, and then of Freud. Janet's classical study of the symptoms of hysteria appeared in 1892 (**33**), and Freud's great book on the interpretation of dreams in 1900 (**24**). Janet's theory of hysteria in terms of dissociation and the retraction of the field of attention was a psychological theory, although not an introspective one. Freud in his association with Breuer discovered the "talking cure" out of which psychoanalysis has emerged (**13**). The effect of psychoanalysis upon psychiatry has during the present century been profound. Not only has psychiatry taken over psychoanalytic concepts while rejecting the total system, but the psychiatric interview has been arranged to assay consciousness, as well as to bring to consciousness those forgotten materials whose absence constitutes a symptom of mental disorder. Nowadays the interview and the couch are used as tools for a special kind of introspection, one which inventories consciousness and seeks to bring forgotten memories up to and across the threshold of introspection.

One of the most definite claims for the use of introspection by abnormal psychology was made by Morton Prince, Janet's complement in America, long a student of dissociated and alternating personalities, and later insistent upon the simultaneous functioning of coconscious personalities (**62, 63**). Prince once suggested that introspections might be obtained simultaneously from two coconscious personalities, even though they had but one set of receptors and effectors between them. You might, he thought, be able to question one personality with written questions shown to the eye and get the protocols spoken by the voice, while the other personality received spoken questions by ear and replied by writing on a pad. This is a difficult form of dissociation and, when it has been tried, the protocols tend to become habituated clichés or nonsense (**69**); yet Prince's suggestion carries the point that patient's protocols are, after all, a kind of introspection. The operationist can, of course, translate protocols into discriminative response, for any consciousness that yields public data can be described in behavioristic terms; yet that fact does not alter the feeling of reality that

the psychopathologists have about both consciousness, got by introspection, and unconsciousness, observed by more inferential technics.

PSYCHOPHYSICS

It was the prevailing nineteenth-century dualism of mind and body, and thus of spiritualism and materialism, that led Fechner, concerned with combating materialism and in establishing a spiritualistic monism, to invent psychophysics (22). By measuring both the physical stimulus and the psychical sensation and by showing how the magnitude of the latter is dependent upon the magnitude of the former, he believed that he was bringing mind and matter into a single system of relationships. The effect of Fechner's success in devising or standardizing the classial psychophysical methods which are still in use was to support the current psychophysical parallelism—although that is not what Fechner intended. For psychophysics the stimulus was available as an independent variable. The sensations, or the relative magnitudes of two sensations, or the sense-distances between two sensations, were available to introspection and so constituted a dependent variable in the psychophysical experiment. This kind of introspection has remained scientifically useful in experimental psychology for a full century and persists in good status today, although of course operationism has the necessary formulas for transforming it into behavioristic terms.

Before Fechner the experimental attack on sensory problems was apt to be psychophysical. Investigators determined both absolute and differential thresholds. When Bouguer in 1760 measured the differential threshold for brightness, he relied on the observer's judgment as to when a shadow on a screen becomes only just noticeable (10, pp. 51 f.). Weber's formulation of his psychophysical law in 1834 depended on the same kind of judgment (92, pp. 44–175). Sensory phenomenology was stimulated by the discovery of the law of the spinal nerve roots (1811, 1822) which showed that the sensory nerves present a set of problems of their own. Johannes Müller's doctrine of specific nerve-energies (1826, 1838) was, in a sense, psychophysics, since it distinguished between sensory quality and the property of the stimulus which arouses the quality (56, pp. 44–55; 57, II, p. v). Many of these early instances of psychophysics, especially the quantitative ones, have been discussed by Titchener (78, II, pt. ii, pp. xiii–cxvi). There is no need to labor the point that parallelism was the accepted doctrine of the century and that psychophysics consisted in the observation of correlations, many of them quantitative, between the two correlated terms of mind and body. No one doubted that you can observe mind as sensory experience.

For at least half a century (1860–1910) psychophysics flourished along with classical introspection and came under some of its constraints. It was thought, for instance, that observers need special training in order to give reliable results. Titchener, as we have already seen, warned against the stimulus-error (5; 79, pp. 202 f.), and both Wundt and Titchener believed that control stimuli (*Vexirversuche*) were improper. For instance, in determining the limen of dual impression upon the skin, you vary the separation of the esthesiometer points according to some standard procedure, but you do not throw in single points as controls—not if you

are a classical introspectionist. The control lies in training the observer to avoid the stimulus-error. If he says *two* when he has only one, he is not wrong, for introspection cannot lie—or at least it was thought that good introspection of trained observers cannot lie very much, and in any case to argue that a one-point stimulus cannot give rise to a two-point perception is to prejudge the experiment which seeks to find what it is that you do feel for every value of the stimulus.

The same point about introspection appears in Wundt's method of identical series for the investigation of recognition (66, pp. 24–30). In this method you give the observer a series of stimulus-objects, and later you give him in the test the identical series again, having him state which items he recognizes. You do not introduce new items as controls. He knows the series are the same, but you trust him in his introspection. He will not report recognition for an item unless he experiences recognition, and no one but the observer himself can publish the privacy of his own consciousness. If you place all this responsibility on the observer, no wonder training becomes important.

This kind of incontrovertible psychophysical introspection did not last long in the functional atmosphere of American psychology. Perhaps it has not now been heard of for thirty years.

For the half century after Fechner the psychophysicists always talked about observing and measuring sensation, but actually they were observing, reporting upon, and measuring, not complete sensations, but sensory attributes. From Fechner on, the psychophysical methods were applied to judgments of the quality, intensity, extensity, or duration of sensory experience, and Külpe, after he broke away from Wundt, suggested that you never actually do observe a whole sensation, but only separately its attributes, out of which you build the sensation up as a scientific construct (42). Later Rahn, a student of Külpe's, reinforced this comment (65), and Titchener ultimately adjusted his views to meet the contention (84).

Külpe in 1893 had argued that the attributes of sensation are (a) inseparable from the sensation (if any attribute becomes zero, the whole sensation ceases to exist) but (b) independently variable with respect to each other (you can change one and keep the other constant) (41, pp. 30–38). Later this view turned out to be wrong, for there are separate attributes, like the pitch and loudness of tones and the hue and brightness of spectral lights, which cannot easily be varied independently by controlling their stimulus. Stevens solved this problem by an appeal to the concept of invariance. You have, he said, an independent attribute if it remains invariant when the dimensions of the stimulus are varied in accordance with some unique determined function (7, 70, 71). This concept results in plotting isesthetic contours on a stimulus diagram, e.g., in plotting isophonic contours for pitch and loudness against stimulus frequency and energy, or isochromatic contours for hue, brightness, and saturation against stimulus wavelength and energy. Sensory equality becomes the crucial datum, but subjective equality is computed from the same basic introspective data that Fechner used—judgments of *greater* and *less* or of some similar complementary categories.

Modern psychophysics is also en-

gaged in the determination of sensory interval scales and ratio scales, and for this purpose observers report on the relation of one sense-distance as greater or less than another (interval scale) or on the ratio of one sensory attribute to another (ratio scale) (75, pp. 23–30). Such introspection is reliable and receives general approval, even in behavioristic America.

There are other less quantitative kinds of psychophysics which still make successful use of reports on sensory experience and which can be properly classified as modern introspection. An excellent example is Crocker's work on the analysis and assessment of flavors by trained panels of judges, persons who are really introspectors especially trained to appreciate and analyze tastes and smells (18). They estimate the degree of the various olfactory and gustatory components in a flavor, check judgments against one another, working as a cooperative team with high motivation and enthusiasm. Such a trained panel may be sent out from the parent laboratory to some industrial plant to savor and calibrate its product, and then may later be brought back to the parent laboratory for checking in introspective reliability and also, when necessary, for analytic recalibration. Crocker's account of how attitudes are fixed and judgments rendered uniform in these panels is reminiscent of the atmosphere of Wundt's laboratory in all respects, except that Crocker's laboratory lacks the authoritarian control of Wundt's.

Another recent example of the modern use of the report of sensory experience is the book on pain by Hardy and his associates (27). This book sets forth the psychophysics of pain, having regard, among other things, to the different qualities of algesic experience, and to establishing a sensory scale of pain by the subjective equation of algesic sense-distance.

The lesson to be learned from psychophysics is, therefore, that, in respect of the observation of sensory experience, introspection has thrived for a hundred years and is still in style.

ANIMAL CONSCIOUSNESS

In denying rational souls to animals, Descartes had made the problem of animal psychology relatively unimportant, but Darwin, with his evolutionary argument that the forms of both mind and body show continuous development from lower species to man (1872), changed all that (19). You began then to hear from Romanes about mental evolution and the evolution of intelligence (1883). Romanes coined the term *comparative psychology* for the study of the nature of mind in different species (67). By giving the animal mind the benefit of the doubt, he was able to represent animal intelligence as not so far below man's. Lloyd Morgan, writing a comparative psychology, sought to temper Romanes' enthusiasm with the principle of parsimony: do not interpret an action as the outcome of the exercise of a higher psychical faculty, he said, if it can be interpreted as the outcome of one that stands lower in the psychological scale (54). Lloyd Morgan warned against "anthropomorphism" in assessing animal behavior—meaning, of course, anthropopsychism. Loeb, establishing the concept of tropism and the unconscious action of lower animal forms (1890), suggested that consciousness emerges in the course of evolution as it becomes needed for more adaptive action and that the faculty of associative memory constitutes a cri-

terion of it (47). Experiments on animal intelligence began, notably Thorndike's in 1898 (77). In the decade 1900–1910 there was marked activity in experimental comparative psychology, a great deal of it concerned with the measurement of animal intelligence for which the maze was regarded as a very useful instrument.

Although there had already been argument put forward in favor of an objective animal psychology (4), comparative psychology got under way in a period when a psychology with consciousness left out was generally regarded as psychology without its psyche—a branch of physiology perhaps. American functional psychology kept consciousness inside the fold, and the comparative psychologists settled on a formula for the observation of animal consciousness which might well have been called *animal introspection*. Nowhere has this problem been more clearly stated than by Washburn in her handbook of 1908 on the animal mind (88, p. 13). She wrote:

If an animal behaves in a certain manner, what may we conclude the consciousness accompanying its behavior to be like? . . . At the outset of our discussion . . . we are obliged to acknowledge that *all psychic interpretation of animal behavior must be on the analogy of human experience.* We do not know the meaning of such terms as perception, pleasure, fear, anger, visual sensation, etc., except as these processes form a part of the contents of our own minds. Whether we will or no, we must be anthropomorphic in the notions we form of what takes place in the mind of an animal.

There is an implication here that you learn about human consciousness by direct observation of it in introspection, but that animal consciousness is known only indirectly by analogical inference. Not every-

one held to that difference, however. Max Meyer put forward what he called the psychology of "the other one," an argument that your own personal consciousness is not material for science, being particular and not general, and that psychology studies always other organisms—other people, other animals (52). In this sense both the animal's conduct and man's words are introspection if they are taken as meaning something about the subject's consciousness. Even Titchener can be found saying of this argument from analogy: "The animal is thus made, so to say, to observe, to introspect; it attends to certain stimuli, and registers its experience by gesture" (79, pp. 30–36).

It is interesting to see how Watson, before he had thought out behaviorism, accepted the current belief of this first decade of experimental animal psychology that knowledge of animal consciousness is the ultimate goal in comparative psychology. Watson was still at Chicago, the home of systematic functional psychology, which held that consciousness is to be understood psychologically in terms of its use to the organism. He had entitled his monograph of 1907: *Kinaesthetic and Organic Sensations: Their Role in the Reactions of the White Rat to the Maze* (89, pp. 90–97). In this investigation he eliminated vision, hearing, taste, smell, and certain cutaneous factors from the repertoire of the rat who still remembers how to run the maze, and he concluded that "intra-organic sensations—the kinaesthetic sensations coupled with the organic probably, and possibly with the static" are what the rat uses in following the correct path. Watson even discussed the possibility of the rat's use of visual imagery, which "in our own case would play a preponderating role." He suggested that success for

the rat as it runs may reassure it: "If the turn is made at the proper stage (and it has been shown that blind rats deprived of their vibrissae can make these turns without allowing their bodies to touch the edges of the openings at the turns), the animal may be supposed thereby to get a 'reassuring feeling' which is exactly comparable to the experience which we get when we touch a familiar object in the dark."

Later, of course, Watson repudiated this supererogatory concern with consciousness and asked psychologists to get closer to their data of stimuli and responses. That was a move toward positivism, but Watson did not think of that. Indeed, it is possible to regard animal behavior as a kind of language which means something about consciousness, just as it is also possible to strip introspection of its meanings and regard it as mere verbal motion. Certainly, if Max Meyer's "other one" can introspect, the animals can too and did before behaviorism made their consciousnesses unimportant.

VERBAL REPORT

Watson's reaction in 1913, away from the pedantry and unreliability of introspection, as he saw it, toward the more positive psychology of stimulus and response, was an attempt, not so much to create behaviorism as a new psychology with consciousness left out, as it was to reformulate the old psychology in new terms (90). For the imagery of thinking, he suggested that we can substitute incipient subvocal movement. Feeling, he believed, might turn out to be endocrine. Association had already been shown by Pavlov so be a conditioning of reflex retponses and not necessarily a connection among ideas. Watson formally ruled introspection out of psy-

chology but he left in the more reliable results of introspection, notably in psychophysics (91). Thus it was necessary for him to leave in introspection as verbal report. Did he thus embrace the bath with the baby? Is introspection anything more than verbal report?

Actually there is a difference. Verbal report viewed simply as behavior is capable of physical specification, in which the writing and speaking of words appear as very different kinds of movements until they have been shown to be equivalent in an experimental situation. On the other hand, verbal report as introspection is not response but observation and description and therefore reference, an indication of objects of observation in the sense of the meanings of the words used.

Another way of expressing this same matter is to write two formulas:

[1] Introspective observation:
$$E \rightarrow O = S \rightarrow \text{facts of consciousness}$$
[2] Behavioristic observation:
$$O = E \rightarrow S \rightarrow \text{facts of psychology}$$

The corresponding sentences are: [1] In introspective observation, the experimenter notes the facts of consciousness which the observer, who is the subject, has observed. [2] In behavioristic observation, the observer, who is the experimenter, observes the behavior of the subject in respect of its implications for the facts of psychology. In classical introspection the subject is the observer. He has responsibility for the correctness of his descriptions of conscious data and thus he had at Leipzig, Cornell, and elsewhere to be trained, for introspection is more than having experience. Behaviorism shifts the locus of scientific responsibility from an observing subject to the experimenter who becomes the observer *of* the subject. In this way it is possible to bring to psychological

observation irresponsible and untrained subjects—animals, children, the feebleminded, the mentally ill, and also the untrained normal human adult. Thus all the mental tests come into psychology because mostly they involve verbal responses from naive subjects. And the animal experiments come in because ordinarily the discriminative behavior of the animals is a language devised by the experimenter and taught to the animal so that he can tell the experimenter about his abilities and capacities. Are we to say that the animal is not introspecting because he is not communicating to himself what he is communicating to the experimenter? Perhaps. The important thing is to see that Watson, in attacking introspection, was objecting, not to the use of words by the subject, but to trusting the subject to use the words only with those meanings that the experimenter wishes the words to have.

INTROSPECTION AS AN OPERATION

Watson, in substituting verbal report for introspection, was moving in the positivistic direction, but the culmination of this movement came later with the acceptance of operational definitions as providing the most secure specification for psychological concepts. Operationism is perhaps a movement toward greater precision in scientific thinking, but it is not a school. American psychologists first picked up this modern form of the old positivism from the physicist, P. W. Bridgman, who was using the technic to explain relativity theory (14). Then it was found that logical positivism, as the movement came to be called later, was developing at the same time among the logicians in Vienna (23, pp. 1–52). Presently it became clear that the two movements were logically the

same. Stevens undertook to be the expositor to American psychologists (74). Bridgman was content to let operational definition go back ultimately to experience, but for psychologists that regression would not do at all. For them experience was a concept in special need of definition, since the availability of consciousness to scientific observation was the main problem dividing the schools (72, 73). The effect of a great deal of discussion along these lines in the 1930's was a change in the status of consciousness from (a) the reservoir of experience upon which all empirical science draws to (b) a concept based upon observation and specified by the observational operations that make conscious data available to science. That is a large change from the introspection that cannot lie because the having of experience is the knowing that you have it.

Nowadays the word *introspection* has dropped out of use. *Consciousness* or *phenomenal experience* or *sensory datum* or some other equivalent mentalistic term indicates a psychological construct which is got by inference *from* the observations. A comparable concept is the *intervening variable*, and a case could be made for Tolman as a phenomenological operationist, directly observing purpose and kindred entities in his data. Do you truly observe consciousness or an intervening variable? Do you observe any construct, or do you infer it? Do you look at the ammeter and observe the strength of the current or is what you observe merely a pointer on a scale?

Thus the answer to the question "What became of introspection?" seems to be this. Introspection as a special technic has gone. The object of introspection—sometimes called *consciousness*, sometimes something else—is a construct like an ability, or

an intervening variable, or a conditioned response, or any of the other "realities" out of which a general psychology is formed. The modern equivalent of introspection persists in the reports of sensory experience in psychophysics, in the protocols of patients with psychological difficulties, in the phenomenological descriptions of perception and other psychological events as provided notably by Gestalt psychologists, and also in a great deal of social psychology and psychological philosophy where the Cartesian dualism is still found to be convenient.

UNCONSCIOUSNESS

Any study of the history of the availability of consciousness to scientific observation, like the present one, gains significance as we consider also the availability of unconsciousness to science. A is specified clearly only with respect to *not-A*. It would not, however, be proper to undertake now the consideration of all the means whereby a knowledge of unconscious psychological events has been brought into science. Nevertheless we may use a paragraph to list the outstanding fields which contributed to what nowadays we call psychology and which got along, nevertheless, without any observation that might be called *introspection*.

The *reflex* was thought almost from its discovery to be unconscious, largely because it could occur without the brain, although Pflüger was of the opinion that its purposiveness implies that it is conscious. Was Lotze, who disagreed with Pflüger, relying on introspection to be sure that reflexes are unconscious? *Instinct* was ordinarily opposed to intelligent action and often supposed to be unconscious. Unconsciousness, however, was not ordinarily involved in its definition; the criterion for instinct was that it was unlearned and usually involuntary. Loeb's *tropism* was defined with consciousness irrelevant. Herbart's *ideas in a state of tendency* were defined as unconscious, as were Fechner's *negative sensations*. Although the Würzburg school was developing systematic introspection, it seems clear now that its great discovery was the existence and effectiveness of unconscious tendencies—the *determining tendency,* the *Aufgabe,* etc. Freud made the concept of the unconscious familiar to everyone and also started the development of the technics of observation that now replace introspection, but the test of unconsciousness (suppression, repression) remained in part introspection, the fact that ideas that might have been expected to be in mind were conspicuously absent. Thus dynamic psychology carries on with the basic assumption that you cannot trust the subject's personal belief (introspection) for the true assessment of his motives.

In all these cases consciousness is seen to have been important in a negative manner, for its absence is a matter of interest and sometimes even an essential specification—as would, indeed, be expected in a psychology that was originally informed on the dualistic pattern. Indeed it is only in a dualism that consciousness has a distinctive meaning.

CONCLUSION

Now let the writer say what he thinks has become of introspection.

There have been in the history of science two important dichotomies that have been made with respect to introspection. (a) The first is animal psychology vs. human psychology: human beings are supposed to be able to introspect, and animals are not. (b) The second is the unconscious

mind *vs.* the conscious mind, with introspection the means of observing consciousness. These two dichotomies reduce, however, to one: inference *vs.* direct experience.

Operational logic, in my opinion, now fuses this single dichotomy because it shows that human consciousness is an inferred construct, a concept as inferential as any of the other psychologists' realities (**32**, p. 184), and that literally immediate observation, the introspection that cannot lie, does not exist. All observation is a process that takes some time and is subject to error in the course of its occurrence.

Introspection's product, consciousness, appears now in the bodies of its progeny: the sensory experience of psychophysics, the phenomenal data of Gestalt psychology, the symbolic processes and intervening variables employed by various behaviorists. the ideas, the manifest wishes, the hallucinations, delusions, and emotions of patients and neurotic subjects, and the many mentalistic concepts which social psychology uses. The newest usage is this latter one, social perception, a term which refers both to the perception of social phenomena, like anger and danger, and the perceptions which are understood by reference to their social determinants; but here the introspection is not different in kind from the phenomenological description that the Gestalt psychologists still use. In general, however, it seems to the writer that there is no longer to be found any sharp dichotomy setting off the introspectable from the unconscious. That once fundamental distinction disappeared with the dissolution of dualism. Consciousness nowadays is simply one of many concepts which psychology employs, usually under some other name, whenever it finds the category useful for the generalization of observations.

REFERENCES

1. ACH, N. *Ueber die Willenstätigkeit und das Denkens.* Göttingen: Vandenhoeck & Ruprecht, 1905.
2. ANGELL, J. R. The province of functional psychology. *Psychol. Rev.,* 1907, 14, 61–91.
3. AVENARIUS, R. *Kritik der reinen Erfahrung.* (2 vols.) Leipzig: Fues & Reisland, 1888–1890.
4. BEER, T., BETHE, A., & VON UEXKÜLL, J. Vorschläge zu einer objektivirender Nomenclatur in der Physiologie der Nervensystems. *Biol. Centbl.,* 1899, 19, 517–521.
5. BORING, E. G. The stimulus-error. *Amer. J. Psychol.,* 1921, **33**, 449–471.
6. BORING, E. G. *The physical dimensions of consciousness.* New York: Century, 1933.
7. BORING, E. G. The relation of the attributes of sensation to the dimensions of the stimulus. *Philos. Sci.,* 1935, 2, 236–245.
8. BORING, E. G. *A history of experimental psychology.* (2nd Ed.) New York: Appleton-Century-Crofts, 1950.
9. BORING, E. G. The influence of evolutionary theory upon American psychological thought. In S. Persons (Ed.), *Evolutionary thought in America.* New Haven: Yale Univer. Press, 1950. Pp. 269–298.
10. BOUGUER, P. *Traité d'optique sur la gradation de la lumière.* Paris: Guerin & Delatour, 1760.
11. BRAID, J. *Neurypnology; or, the rationale of nervous sleep; considered in relation with animal magnetism.* London: Churchill, 1843.
12. BRENTANO, F. *Psychologie vom empirischen Standpunkte.* Leipzig: Duncker & Humblot, 1874.
13. BREUER, J., & FREUD, S. *Studien über Hysterie.* Leipzig & Vienna: Deuticke, 1895.
14. BRIDGMAN, P. W. *The logic of modern physics.* New York: Macmillan, 1927.
15. CHARCOT, J. M. *Oeuvres complètes.* (9 vols.) Paris: Bur. Prog. méd., 1886–90.
16. CLARKE, H. M. Conscious attitudes. *Amer. J. Psychol.,* 1911, 22, 214–249.

17. COMTE, A. *Cours de philosophie positive.* Paris: Bachelier, 1830–42. (Cf. Nos. 15, 57.)

18. CROCKER, E. C. *Flavor.* New York: McGraw-Hill, 1945.

19. DARWIN, C. *Expression of the emotions in man and animals.* London: Murray, 1872.

20. DESCARTES, R. *Les passions de l'âme.* Paris: Le Gras, 1649.

21. EISLER, R. *Wörterbuch der philosophischen Begriffe.* Berlin: Mittler & Sohn, 1910.

22. FECHNER, G. T. *Elemente der Psychophysik.* Leipzig: Breitkopf & Härtel, 1860.

23. FRANK, P. *Modern science and its philosophy.* Cambridge, Mass.: Harvard Univer. Press, 1949.

24. FREUD, S. *Die Traumdeutung.* Leipzig: Deuticke, 1900.

25. GEISSLER, L. R. The measurement of attention. *Amer. J. Psychol.*, 1909, 20, 473–529.

26. GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.

27. HARDY, J. D., WOLFF, H. G., & GOODELL, H. *Pain sensations and reactions.* Baltimore: Williams & Wilkins, 1952.

28. HAYES, S. P. A study of the affective qualities. *Amer. J. Psychol.*, 1906, 17, 358–393.

29. HUSSERL, E. G. *Logische Untersuchungen: Untersuchungen zur Phänomenologie und Theorie der Erkenntnis.* Halle: Niemeyer, 1901.

30. HUXLEY, T. H. On the hypothesis that animals are automata, and its history. *Fortnightly Rev.*, 1874, 22 (N.S. 16), 555–580.

31. JACOBSON, E. On meaning and understanding. *Amer. J. Psychol.*, 1911, 22, 553–577.

32. JAMES, W. *Principles of psychology.* New York: Holt, 1890.

33. JANET, P. *L'état mental des hystériques.* Paris: Rueff, 1892.

34. KATZ, D. Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung. *Zsch. Psychol.*, 1911, Ergbd. 7. Leipzig: Barth, 1911.

35. KÖHLER, W. Intelligenzprüfung an Anthropoiden. *Abhl. preuss. Akad. Wiss. Berlin (phys.-math. Kl.),* 1917, nr. 1.

36. KÖHLER, W. *Die physische Gestalten in Ruhe und im stationären Zustand.* Braunschweig: Vieweg & Sohn, 1920.

37. KÖHLER, W. *Gestalt psychology.* New York: Liveright, 1929.

38. KOFFKA, K. Perception: an introduction to Gestalt-theorie. *Psychol. Bull.*, 1922, 19, 531–585.

39. KOFFKA, K. *Principles of Gestalt psychology.* New York: Harcourt Brace, 1935.

40. KRAEPELIN, E. *Psychiatrie.* (5th Ed.) Leipzig: Barth, 1896.

41. KÜLPE, O. *Grundriss der Psychologie.* Leipzig: Engelmann, 1893.

42. KÜLPE, O. Versuche über Abstraktion. *Ber. I. Kongr. exper. Psychol.*, 1904, 56–68.

43. KÜLPE, O. *Vorlesungen über Psychologie.* Leipzig: Hirzel, 1920. (Posthumous)

44. LA METTRIE, J. O. *L'homme machine.* Leiden: Luzac, 1748.

45. LIÉBEAULT, A. A. *Du sommeil et des états analogues, considerérés surtout au point de vue de l'action de la morale sur le physique.* Paris: Masson, 1866.

46. LOCKE, J. *Essay concerning human understanding.* London: Basset, 1690.

47. LOEB, J. *Der Heliotropismus der Thiere und seiner Ueberstimmung mit dem Heliotropismus der Pflanzen.* Würzburg: Hertz, 1890.

48. MACH, E. *Beiträge zur Analyse der Empfindungen.* Jena: Fischer, 1886.

49. MARBE, K. *Experimentell-psychologische Untersuchungen über das Urteil.* Leipzig: Engelmann, 1901.

50. MAYER, A., & ORTH, J. Zur qualitativen Untersuchung der Association. *Zsch. Psychol.*, 1901, 26, 1–13.

51. McDOUGALL, W. *Outline of psychology.* New York: Scribner's, 1923.

52. MEYER, M. *Psychology of the other one.* Columbia, Mo.: Mo. Book Co., 1921.

53. MILL, J. S. *Auguste Comte and positivism.* (3rd Ed.) London: Trübner, 1882.

54. MORGAN, C. L. *Introduction to comparative psychology.* London: Scott, 1894.

55. MÜLLER, G. E. Zur Analyse der Gedächnistätigkeit und des Vorstellungsverlaufes, I. *Zsch. Psychol.*, Ergbd. Leipzig: Barth, 1911.

56. MÜLLER, J. *Zur vergleichenden Physiologie des Gesichtssinnes.* Leipzig: Cnobloch, 1826.

57. MÜLLER, J. *Handbuch der Physiologie des Menschen.* (3 vols.) Coblenz: Hölscher, 1833–40.

58. NAKASHIMA, T. Contributions to the study of the affective processes. *Amer. J. Psychol.*, 1909, 20, 157–193.

59. OKABE, T. An experimental study of belief. *Amer. J. Psychol.*, 1910, 21, 563–596.

60. ORTH, J. *Gefühl und Bewusstseinslage.* Berlin: Reuther & Reichard, 1903.

61. PINEL, P. *Traité médico-philosophique*

aliénation mentale. Paris: Richard, Caille & Revier, 1801.

62. PRINCE, M. *The dissociation of a personality.* New York: Longmans Green, 1905.

63. PRINCE, M. *The unconscious.* New York: Macmillan, 1914.

64. PYLE, W. H. An experimental study of expectation. *Amer. J. Psychol.*, 1909, 20, 530–569.

65. RAHN, C. The relation of sensation to other categories in contemporary psychology. *Psychol. Monogr.*, 1914, 16, No. 1 (Whole No. 67).

66. REUTHER, F. Beiträge zur Gedächtnisforschung. *Psychol. Stud.*, 1905, 1, 4–101.

67. ROMANES, G. J. *Mental evolution in animals.* London: Kegan, Paul, Trench, 1883.

68. RUBIN, E. *Synsoplevede figurer.* Copenhagen: Gyldendal, 1915.

69. SOLOMONS, L. M., & STEIN, G. Normal motor automatism. *Psychol. Rev.*, 1896, 3, 492–512.

70. STEVENS, S. S. The attributes of tones. *Proc. nat. Acad. Sci., Wash.*, 1934, 20, 457–459.

71. STEVENS, S. S. Volume and intensity of tones. *Amer. J. Psychol.*, 1934, 46, 397–408.

72. STEVENS, S. S. The operational basis of psychology. *Amer. J. Psychol.*, 1935, 47, 323–330.

73. STEVENS, S. S. The operational definition of psychological concepts. *Psychol. Rev.*, 1935, 42, 517–527.

74. STEVENS, S. S. Psychology and the science of science. *Psychol. Bull.*, 1939, 36, 221–263.

75. STEVENS, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 7–49.

76. STUMPF, C. Erscheinungen und psychische Funktionen. *Abhl. pruess. Akad. Wiss. Berlin (philos.-hist. Kl.)*, 1906, nr. 4.

77. THORNDIKE, E. L. Animal intelligence. *Psychol. Monogr.*, 1898, No. 2 (Whole No. 8).

78. TITCHENER, E. B. *Experimental psychology.* (2 vols., 4 pts.) New York: Macmillan, 1901–05.

79. TITCHENER, E. B. *A text-book of psychology.* New York: Macmillan, 1910.

80. TITCHENER, E. B. Description vs. statement of meaning. *Amer. J. Psychol.*, 1912, 23, 165–182.

81. TITCHENER, E. B. Prolegomena to a study of introspection. *Amer. J. Psychol.*, 1912, 23, 427–448.

82. TITCHENER, E. B. The schema of introspection. *Amer. J. Psychol.*, 1912, 23, 485–508.

83. TITCHENER, E. B. *A beginner's psychology.* New York: Macmillan, 1915.

84. TITCHENER, E. B. Sensation and system. *Amer. J. Psychol.*, 1915, 26, 258–267.

85. TITCHENER, E. B. *Systematic psychology: prolegomena.* New York: Macmillan, 1929. (Posthumous)

86. TOLMAN, E. C. Operational behaviorism and current trends in psychology. *Proc. 25th Anniv. Celbr. Inaug. Grad. Stud. Univer. So. Calif.* Los Angeles: Univer. So. Calif. Press., 1936, 89–103.

87. WARD, J. *Psychological principles.* Cambridge, Eng.: University Press, 1918.

88. WASHBURN, M. F. *The animal mind.* New York: Macmillan, 1908.

89. WATSON, J. B. Kinaesthetic and organic sensations: their role in the reactions of the white rat to the maze. *Psychol. Monogr.*, 1907, 8, No. 2 (Whole No. 33).

90. WATSON, J. B. Psychology as the behaviorist views it. *Psychol. Rev.*, 1913, 20, 158–177.

91. WATSON, J. B. *Psychology from the standpoint of a behaviorist.* Philadelphia: Lippincott, 1919.

92. WATT, H. J. Experimentelle Beiträge zur einer Theorie des Denkens. *Arch. ges. Psychol.*, 1905, 4, 289–436.

93. WEBER, E. H. *De pulsu, resorptione, auditu et tactu: annotationes anatomicae et physiologicae.* Leipzig: Koehler, 1834.

94. WERTHEIMER, M. Experimentelle Studien über das Sehen von Bewegungen. *Zsch. Psychol.*, 1912, 61, 161–265.

95. WERTHEIMER, M. Untersuchungen zur Lehre von der Gestalt. *Psychol. Forsch.*, 1921, 1, 47–58; 1923, 4, 301–350.

96. WHITE, R. W. *The abnormal personality.* New York: Ronald, 1948.

97. WUNDT, W. Selbstbeobachtung und innere Wahrnehmung. *Philos. Stud.*, 1888, 4, 292–309.

98. WUNDT, W. *Grundriss der Psychologie.* Leipzig: Engelmann, 1896.

99. ZILBOORG, G. *A history of medical psychology.* New York: Norton, 1941.

A BRIEF CRITICAL REVIEW OF LOUDNESS RECRUITMENT

J. DONALD HARRIS

Medical Research Laboratory
Submarine Base, New London, Connecticut

The growth of loudness in the normal ear when the energy of a pure tone is increased above the absolute intensive threshold has been well specified by workers too numerous to mention here. In less detail, the growth of loudness in the normal ear has been studied in the cases of speech and of noise. For the ear which is not normal, however, there are a variety of aberrations in loudness which may occur during an increase in signal strength. Chief of these is that, for an equivalent increase in signal strength, the growth of loudness in a non-normal ear may outstrip by far the growth of loudness in a normal ear. By the time a tone is raised 100 db over normal threshold, a partial deafness of up to 50 or 60 db may be rendered negligible. This paradox, revealed by a variety of experimental and clinical techniques, was named "recruitment" by Fowler. The term is now commonly used in this country, though the terms regression, recuperation, *lautstarkeäusgleich* (i.e., loudness-compensation), and Fowler-phenomenon are more common in other countries.

This general phenomenon which has its behavioral manifestations in a wide variety of psychoacoustic situations has importance for studies of the physiology and neurology of loudness. It is primarily, however, a psychological manifestation, and it is rather surprising that to date no original article has appeared on this topic in the psychological journals.

The practical importance of recruitment is rapidly claiming the attention of a great number of clinics. Audiometer manufacturers in this country and abroad are designing and selling equipment specifically for the purpose of studying recruitment. It is now widely used as an aid in fixing the locus of an auditory impairment as between the middle and the inner ear, and has even been suggested as further distinguishing the locus as between the organ of Corti and the auditory nerve.

Theoretically, the importance of recruitment lies in what it can tell us of the nature of loudness in the normal-hearing ear. It is now clear that the neurological explanation of recruitment advanced by Lorente de Nò (23) and the explanation in terms of constant loudness-loss (112) (both views receiving the widest circulation) do not account for the clinical data, and so can contribute nothing to the understanding of normal loudness phenomena. Yet, if we understood how, in recruiting ears, the loudness function can take a variety of quite different shapes, we would better understand the normal case. The experimental facts, not always consistent among themselves, are now quite voluminous, and a critical review of recruitment seems appropriate at this time before the now rapidly oncoming avalanche of clinical data gives an illusion of complete knowledge, burying any attempt to arrive at a more fundamental knowledge of the phenomenon.

190

History

Haberman (40) was among the first of modern otologists to note that certain individuals understood the spoken better than the whispered voice. This is true in some recruiting ears. Gradenigo (35) devised the phrase *Index Vocalis* for the quotient of the distance at which the whispered voice is just audible, divided by the same measure for the spoken voice. In a recruiting ear the distance for the spoken voice is relatively large, and the *Index Vocalis* is smaller. The *Index Vocalis* has been determined for the normal ear for a variety of speech sounds by Gradenigo, Zwaardemaker (128, 129), and Hiddema (49). Veis (121) first applied the Index to clinical material, finding a smaller value than normal with some types of hearing loss.

Pohlman and Kranz (100) first used pure tones rather than speech in recruiting ears. They found that partial tonal gaps in the audiogram seemed to disappear at suprathreshold intensities. The method they used was a forerunner of the quantitative method of monaural loudness balancing between two frequencies, to one of which the ear is relatively insensitive.

The clinical possibilities of recruitment were first clearly seen by Fowler (21, 22, 23, 24, 25, 26, 27, 28, 29, 30), who devised a number of techniques for collecting and reporting the data. As a result largely of his articles, every up-to-date otologist has by now added to his armamentarium some test for recruitment. Fowler's general conclusion was that recruitment is an index of nerve deafness, and forms a valuable tool to diagnose that condition.

Katz and Salis (64) first carefully studied the shape of the speech articulation function in the recruiting ear. From later extensions of this important work it is possible to attack the question of what the fundamental auditory abilities are which contribute to understanding speech. Thus, the garbling of speech by a recruiting ear may perhaps eventually be explained partly by a shift in pitch cues, partly by a shift in intensity cues, and partly by a change in the temporal pattern of the loudness of steady-state sounds.

The behavior in noise of normal and of recruiting ears was investigated by Langenbeck (68, 70), who showed distinct differential effects. These results have been the basis for a wide variety of tests of masking as it affects reception for pure tones and for speech (7, 8, 36, 37, 38, 39, 56, 61, 62, 93, 123, 125, 126).

Bustamente Gurria and Garibay (9) made a significant contribution when they separated disorders of the organ of Corti from deafness due to a failure of auditory nerve conduction. They point out that deafness due to impossibility of movement of endolymph, in certain conditions of the labyrinthine capsule, should theoretically be classed as conduction deafness, since it is for mechanical reasons that the organ of Corti cannot function properly.

De Bruïne-Altes (7) published the most extensive monograph to date on recruitment. She presented full case histories on 47 partially deaf patients, with complete data on recruitment collected by whichever techniques were appropriate for the particular patient. A wide variety of types of deafness was included, and the monograph is a storehouse of valuable information. Unfortunately, no patients were available with tumor of the VIII nerve. This theoretically very important group would

have added much to de Bruïne-Altes' thinking.

Dix, *et al.* (15) studied 20 cases of verified degeneration of the auditory nervous system, most of which showed no recruitment at all. This study and later work (17, 106) showed the need to re-evaluate the general agreement with Fowler's and Lorente de Nò's conclusion that recruitment must be "pathognomonic" of nerve involvement. It may be confined to non-neural structures in the cochlea (80, 96, 127).

The first indication that the condition underlying recruitment might affect other psychoacoustic relationships besides the loudness function (and related phenomena), was given by Lüscher and Zwislocki (83), who showed that in a recruiting ear it took longer, in milliseconds, for a pure tone to attain a "pitch-quality." Since then, several other facts have appeared relating recruitment to other nonloudness phenomena (12, 48, 74, 91, 109).

Our present view, then, is that in recruitment we have an aberration of loudness which has no good explanation (though it can indeed be ascribed to the sense organ rather than to the nerve), which intrudes into many aspects of hearing (and not only those aspects associated with loudness), and which is extremely valuable in the clinic as a diagnostic sign.

PREVIOUS REVIEWS

Important general discussions of recruitment are given (33, 43, 45, 50, 96, 114, 119), especially of its clinical significance (7, 9, 16, 17, 22, 23, 30, 31, 55, 56, 57, 59, 78, 82, 84, 122, 123, 127). Fairly complete instructions for determining the presence of recruitment by two or more methods are provided (7, 13, 22, 31, 62, 65, 81, 88, 102, 120, 122, 123).

PSYCHOACOUSTIC SITUATIONS AFFECTED BY RECRUITMENT

In general, the presence of recruitment in an ear will affect all those psychoacoustic relationships connected with loudness, but it is also known to affect such relationships as masking, fatigue, pitch perception, and the intelligibility of speech.

The Loudness Function

Recruitment, defined as an abnormally rapid growth of loudness, should perhaps best be determined in the suspected ear by actually deriving the loudness function for that ear from the usual processes of fractionation, bisection, etc. Surprisingly enough, no one has published a loudness function, derived in this fundamental way, for a partially deafened ear. This is probably because those individuals who contributed to the definition of the normal loudness function, mostly classical psychologists and communications engineers, have not usually had a corresponding interest in clinical deafness.

It would be instructive to derive the loudness function by the usual means in a series of partially deafened ears, and to compare individual cases with the normal function. The objection could be raised that the comparison is impossible because there would be no way to equate a loudness of one sone between two quite different individuals. This is true. However, something could be done by studying the slope of the loudness function, or by selecting individuals with one normal ear and one ear partially deafened and recruiting. Loudness functions drawn from these two ears could be drawn on the same coordinates, the two curves coinciding at some point, say at one sone,

found by loudness balancing to be equally loud.

A short-cut substitute for the rather laborious task of drawing independent loudness functions for the two ears of a monaurally, partially deaf patient was proposed by Fowler (21) and followed by practically all subsequent workers in this field. It was assumed that the one ear would produce a normal loudness function if the usual procedures were applied; the suspected ear could then be matched to it at a succession of loudnesses. Because by far the greater part of this work has been done with the clinical audiometer, the unit used has been Sensation Level (in db over normal threshold on the audiometer). What is actually found, then, is difference in db between threshold for the two ears, and the lessening of this difference at equal loudness at higher sensation levels. It is of course possible to re-draw the data in true loudness units, converting sensation level to loudness level (114, p. 123), and thence to loudness in sones (34). But in practice this is never done.

It is important to emphasize that it is possible to interpret a loudness function from a recruiting ear only if a perfectly normal ear exists for comparison. Much of the work on ramifications of recruitment has resulted from a search for an index of recruitment where neither ear can be considered normal.

Isophonic Contours

Where it is difficult to interpret a loudness function from a suspected ear because the other ear is not normal for that frequency, it is sometimes possible to confine one's efforts solely to the suspected ear if that ear is normal for *some* frequency. Monaural loudness matches can then be

obtained between the two frequencies, and an interpretable loudness function drawn for the suspected ear. This has never been reported, but again a short-cut in terms of sensation level has been proposed (7) and widely used. A modification, the drawing of isophonic contours in a suspected ear, has been extensively used (11, 70, 71, 100, 104, 110) by comparing the contours with the normal case (19). It often happens in a partially deafened but recruiting ear that at threshold (isophonic contour = 0) quite a loss in loudness is seen over some range, but that at higher intensities the contours do not differ appreciably from the normal. A full set of isophonic contours is, indeed, about the only feasible way to locate the presence and extent of recruitment over the whole auditory area.

Restriction of Useful Intensity Range

As is well known, the threshold of discomfort for pure tones is about 120 db sound pressure level (SPL). Yet this threshold may not change for a partially deafened but recruiting ear. Thus, if a recruiting deafness of 60 db exists, that ear may have a dynamic range of only 60 db between audibility and discomfort. The striking difference between this and the normal case is apparent.

Intensity Discrimination

As a corollary of crowding several hundred thousand millisones into a 60-db span, as explained above, each db should encompass more millisones than normally, and intensity discrimination should improve. An improved intensive DL was proposed as an index of recruitment (6, 22, 89) and has been used in several clinics (14, 75, 80, 82, 84, 85, 86, 88, 97, 98, 99). A difference may exist be-

tween data gathered by the method of constant stimulus differences (14, 75) and by amplitude modulation (80, 82, 84, 85, 86, 88, 97, 98, 99).

Precision of Absolute Intensive Limen

As another corollary of a restricted dynamic intensity range, each db at or near absolute threshold should contribute more to the clarity of the distinction between "tone present" and "tone absent." Fowler (23) proposed the interval of uncertainty of the psychometric function, as an index of recruitment, but only a rough approximation is in use with audiometers which automatically record the db range through which the patient continuously hunts his threshold (2, 3, 4, 5, 42, 107, 108). This range is often miscalled an intensive DL.

Auditory Fatigue

The question of the relationship between recruitment and auditory fatigue, as opened up by de Maré (89), has proved very fruitful in both areas. Assessing the fatigability of an ear has been proposed for differentiating perceptive-deaf (and therefore, presumably, recruiting) ears (32, 54, 58, 87, 90). On the other hand, the presence of recruitment in a fatigued ear (1, 10, 44, 45, 72, 73) may give an idea of the locus of the fatigue effect within the auditory system (46).

It has been supposed (87, 91) that in a recruiting ear a stimulating tone has a more far-reaching effect up and down the basilar membrane. In these experiments, however, loudness of stimulation was not equated between the normal and the recruiting ear.

Masking

The presence of recruitment in-

creases the masking effect of pure tones (7, 8, 36, 37, 38, 39) and also of noise (56, 61, 69, 92, 93, 125, 126).

Speech Reception

The relation between hearing for weak vs. loud speech is discussed (7, 16, 17, 22, 27, 28, 29, 31, 44, 45, 48, 59, 60, 64, 95, 112, 113, 124, 127). A distinction must be made between the loudness of speech and its intelligibility (31, 43, 45, 101, 113). The shape of the speech articulation curve in recruitment is often quite different from normal (11, 16, 17, 22, 27, 28, 29, 45, 51, 59, 60, 63, 64, 113, 124). Two possible explanations are given (27, 51).

De Maré (91) has suggested that pitch discrimination may be affected by recruitment and may form the basis of a clinical test. Two workers (48, 109) have taken this suggestion.

Pupillary dilation upon high-frequency stimulation was thought to be accentuated in recruitment (74).

In a recruiting ear a longer time than usual is necessary for a short tone to achieve a true tonal quality (83).

CHARACTERISTICS OF RECRUITMENT

Inducing Conditions

Reports that "nerve deafness" or "perceptive deafness" is accompanied by recruitment are too numerous to cite, but we now understand that these terms are too general, early workers using them indiscriminately to refer to any deafness localized in the cochlea. There are, however, a number of case histories in the literature from which we may ascertain rather specifically the cochlear conditions giving rise to recruitment.

Boilermaker's disease, which is always accompanied by recruitment (40), is now understood to be the

result of acoustic trauma whose first manifestation is the bruising of the organ of Corti, and its final sloughing off from the basilar membrane into the endolymph of the scala media. Many cases of Ménière's disease have been proved to have no nerve involvement, the effect residing solely in the sense organ (7, 12, 15, 17, 20, 47, 80, 82), and almost all of these cases exhibit strong recruitment. All 38 cases of "traumatic perceptive deafness," and all 16 cases of "progressive hereditary perceptive deafness" in Bruïne-Altes' series (7) exhibited recruitment. On the other hand, all cochlear damage need not yield recruitment. It is absent in presbyacusis (7, 83, 91) and streptomycin damage (80), and may be absent in basal skull fracture (7). We have seen that true nerve deafness does not as a rule produce recruitment (7, 15, 16, 17, 80, 82, 105, 106, 127). Finally, psychogenic deafness may produce the opposite of recruitment (79), namely, an *increase* in the intensity DL by a factor of 2–3 over the normal.

In general, the facts seem to be consistent with the view of Mygind (96), who argues that recruitment is the result of conductive impairment within the cochlea. It is clear, of course, that a variety of types of impairment can occur, and we may perhaps expect recruitment data to show variations according to type of impairment.

Shapes of Recruitment Curve

Practically all our information concerning the precise shape of the loudness function in recruiting ears comes from studies of binaural loudness balancing, one ear being normal. Data have generally been plotted in terms of audiometer units rather than in loudness units, the ordinates representing db above normal threshold. It is clear that with two normal ears, a plot of sensation levels to achieve equal loudness will be a 45° diagonal. If, however, one ear is defective 20 db at threshold, but recruits completely within 30 db, then the plots will fall *off* the diagonal by 20 db at zero loudness, but will fall *on* the diagonal at the 50 db point. The question for the moment is, what is the exact shape of the curve by which we pass from 20 db *off* the diagonal, to the diagonal. Reger (103, 104) found the slope to increase with extent of deafness, and that the slope was steepest just above the deafened threshold. Steinberg (110) reasoned that the curve should approach the diagonal asymptotically. In general, the shape of the recruitment curve has received less attention than it deserves. Some types of disorders, e.g., those due to long-lasting auditory fatigue, produce plots on a straight line between the point for zero loudness and the point on the diagonal for complete recruitment, and others, e.g., Ménière's disease and simulated deafness by noise-masking, produce an asymptotic curve (45). It sometimes happens that a curve lies asymptotic not to the 45° diagonal, but to a line parallel with the diagonal but displaced some decibels. It has been supposed that this displacement represents a component of conductive (i.e., nonrecruiting) deafness (112, 113). However, if a conductive component is artificially added to an asymptotically recruiting ear by using an earplug, the conductive impairment manifests itself in an initial "delay" *before* the curve tends toward the diagonal. Again, this "delay" is commonly seen as a parallelism extending through 10–15 db before the curve approaches the diagonal in some cases

of inner-ear deafness with no conductive component (the comparison ear being perfectly normal), and in some cases of inner-ear deafness simulated by long-lasting auditory fatigue (never in simulated deafness by noise masking).

Considerable data must still be gathered in an attempt to discover just what conditions within the cochlea will yield certain types of recruitment curve.

Temporal Pattern

An important aspect of recruitment is that it may operate only during the first half-minute or so after stimulus onset. By using simultaneous binaural loudness balance, Hood (52, 53) and Hallpike and Hood (41) have charted a marked decline in loudness for a recruiting ear, beginning at stimulus onset and reaching a stable level only after a minute or less. A normal or a conductive-deaf ear shows no such decline in loudness under comparable conditions. This effect can completely change the interpretation of an ear from recruiting to nonrecruiting. Hood thinks it likely that here is further evidence that recruitment is confined to hair cells, and specifically their on-effect.

It is possible that another expression of the on-effect of Hood is found in the observation by Kobrak (65, 66, 67) that in perceptive deafness the patient experiences a "shock" at the initial attack of a tuning fork.

Overrecruitment

It is amazing enough that complete recruitment appears, such that a threshold deficit in one ear may be completely overcome at some higher intensity, but we find in some cases that when still higher intensities are explored, the initial deficit is more

than overcome—the same physical intensity sounds appreciably louder in the defective than in the normal ear (7, 12, 15, 17, 18, 23, 47). There is, however, a limit to which over-recruitment goes—even at very high intensities it usually amounts to only a few db. No full explanation has ever been given for the phenomenon (Lorente de Nò's explanation of over-recruitment in terms of central inhibition is the only guess attempted to date) though undoubtedly one would be forthcoming if we had a good explanation of recruitment itself; there is, after all, no reason why an abnormally rapid growth of loudness should cease just when equality with a normal ear is achieved. The same compression of loudness units could as well go on at even high intensities.

THEORIES

In the absence of knowledge of just what derangements within the cochlea induce recruitment, and of just how these derangements affect loudness, a number of theories have been offered to account for the facts.

The Fiber-Loss Theory

Steinberg's (110) explanation of recruitment was that if a few nerve fibers were defective this would have a smaller and smaller effect as the stimulus intensity increased. This explanation makes the explicit assumption that loudness is some function of number of active fibers.

The Occlusion Theory

In a discussion of a paper by Fowler (23) Lorente de Nò presented an explanation of recruitment: "If a number of hair cells in the ear or a number of fibers in the cochlear nerve is missing, the tones will appear to be weaker in intensity when near-threshold stimuli are used; but if the

intensity of the tone is increased, the more strongly activated hair cells or cochlear fibers will be sufficient to saturate, i.e., to excite the limiting intensity of the cochlear fiber or the cells of the cochlear nuclei, so that the cerebral cortex will receive the same number of impulses per second for both ears and will perceive the tone delivered to the diseased ear as strongly as the tone delivered to the normal or less affected ear" (p. 220).

It will be noted that this theory applies to both hair cell *and* nerve cell damage, and as such is not affected by later data on recruitment in nerve deafness.

The Constant Loudness-Loss Theory

Steinberg and Gardner (112) and Steinberg (111) attempted an explanation in psychological terms. Suppose an ear to be 20 db deaf for a certain frequency. A tone at threshold for this particular ear amounts to a loudness of about 100 loudness units for the normal ear. But suppose the tone were increased 20 db over the deafened threshold. A tone of this intensity amounts to about 1,000 loudness units for the normal ear. A further increase of 20 db intensity produces 4,500 loudness units, and so on. It is easily seen that because of the nature of the relationship between sensation level and loudness, an ear with a type of deafness resulting in a constant loudness loss would tend to overcome this handicap at high intensities, where the *per cent* loudness loss could be unnoticeable. (In the above illustration, at 40 db sensation level per cent loudness loss is only about 2.0.) This *constant loudness-loss theory* has been given wide currency (18, 50, 114).

However, when the theory is put to a rigorous test (45) it appears quite inadequate. Most cases of asymptotic and straight-line recruitment exhibit a *continuous increase* in loudness loss, rather than a constancy. When the facts were ascertained on a patient from Dr. Fowler's practice (22), a patient often used to buttress the constant loudness-loss theory, it was found that the loudness loss at threshold (4,350 millisones) had risen within 10 db to a loss of 6,975 millisones, and within another 10 db to a loss of 11,045 millisones (44).

The Duplicity Theory

Lurie (77) explained recruitment in terms of the difference in sensitivity of the outer and inner hair cells. If the more sensitive outer hair cells are defective, then the threshold would be raised; but if the sound intensity were raised sufficiently to stimulate the inner hair cells, these would respond normally and a rather sudden increase of loudness might result. This explanation makes the implicit assumption, however, that in the normal ear the outer hair cells no longer contribute significantly to loudness at intensities which stimulate the inner hair cells. It is fairly clear that such is not the case. However, the division of effort among the inner and outer hair cells rests upon good histological (94) and other evidence (77). If these lines of thought are correct, it would indeed be surprising if no aberrations of loudness resulted from some condition which reduces the output of the outer hair cells (as some acoustic trauma seems certain to do) and restricts functioning to the more simply innervated inner hair cells.

The Geometric Theory

Tumarkin (117, 118) proposed that if only the more sensitive hair cells are damaged, then at some higher

intensity the inner hair cells will begin to function, and to these units the brain assigns a certain loudness. So far he follows Lurie's notions. But it is assumed that a sound, which has a certain psychological loudness to a normal ear, has a characteristic geometrical pattern of activity on the basilar membrane. Now if some hair cells in that ear are damaged, the resultant pattern for that sound will be affected, but the missing features will be "filled in" by the "memory," and the psychological contribution to loudness will override the physical deficiency. While this theory is not easily open to experimental test, Tumarkin feels supported by recent evidence that recruitment may be confined to the organ of Corti.

The Impedance Theory

Veckmans (119, 120) attempts an explanation of recruitment in physical terms. Substituting in Ohm's Law, loudness = energy/impedance. Now if the impedance changes in a particular ear as a function of sound intensity, then the usual relation between loudness and energy would be changed, and recruitment might result. Veckmans points out that impedance does change (as a function of frequency) in the normal ear, so that loudness can be said to recruit at 100 cps as compared with, for example, 1,000 cps. And he insists, without experimental evidence, that if an abnormal condition existed which caused impedance to increase with increased energy, then recruitment could be expected. The impedance theory, moreover, is contradicted by the facts of loudness balance between air and bone conduction on the same ear (45). Being of greater impedance, the bony channel should exhibit recruitment, but does not.

The Microphonic Theory

Dix, et al. (15) gives an explanation by analogy, referring to the case of a carbon microphone defective for low intensities but satisfactory for high. This analogy advances our thinking only very slightly.

Any theory of recruitment must meet three demands: it must explain how loudness can grow faster than normally, it must explain overrecruitment and the on-effect, and it must not rely upon neural lesion. The fiber-loss theory, and the constant loudness-loss theory, cannot explain overrecruitment. The geometric theory cannot well be tested. The impedance theory is contradicted upon other grounds. The microphone theory is only an analogy. The duplicity theory can explain overrecruitment in the same terms of relaxing of inhibition as stated by Lorente de Nò, but it cannot explain the case of quick recruitment in a severely deafened ear, where all the more sensitive outer hair cells may presumably be changed. One or more of these theories may apply in a particular case. There is of course no necessity for a single theory to cover all cases of recruitment, since, as we have seen, a wide variety of conditions can give rise to it. The occlusion theory of Fowler and Lorente de Nò, however, still seems adequate to meet all demands. It is only necessary to emphasize that in this theory no neural damage is necessary, the effect being the same if certain sense-organ cells are defective.

ANALOGIES IN THE NORMAL EAR

Békésy (1) used alternate binaural loudness balance with partial deafness simulated by residual fatigue. Later workers have corroborated his pseudo-recruitment (10, 45, 46, 72).

The same technique with noise-masking to simulate degrees of deafness shows quick recruitment (45, 112, 113, 115, 116). The rate of growth of loudness in low tones vs. high tones has been cited as an analogue of recruitment (45, 119). An analogy in the cochlear microphonic has been cited (76). The direct exploration of recruitment in the animal by a behavioral technique has yet to be worked out. Several techniques used in the clinic are adaptable to the animal as well, however, and when this is accomplished one may hope to see the physiology of recruitment illuminated more clearly than at present.

REFERENCES

1. BÉKÉSY, G. v. Zur Theorie des Hörens. Über die eben merkbare Amplituden- und Frequenzänderung eines Tones. Die Theorie der Schwebungen. *Phys. Z.*, 1929, 30, 721–745. Summarized in: *Ann. de Psychol.*, 1930, 31, 63–96.

2. BÉKÉSY, G. v. Über ein neues Audiometer. *Arch. Elekt. Übertragung*, 1947, 1, 13–16.

3. BÉKÉSY, G. v. A new audiometer. *Acta Oto-laryng.*, Stockh., 1947, 35, 411–422.

4. BÉKÉSY, G. v. The recruitment phenomenon and difference limen in hearing and vibration sense. *Laryngoscope*, St. Louis, 1947, 57, 765–777.

5. BÉKÉSY, G. v. Über die recruitmentmessung beim Hören und beim Vibrationssinn mittels eines neuen audiometers. *Acta Oto-laryng.*, Stockh., 1948, Suppl. 74, 26–28.

6. BRINITZER, W. Die Unterschiedsempfindlichkeit für Lautstärken bei Gehörerkrankungen. Versuche mit dem Otoaudion. *Mschr. Ohrenheilk.*, 1935, 69, 1301–1321.

7. BRUÏNE-ALTES, JOHANNA C., DE. *The symptom of regression in different kinds of deafness.* Groningen: J. B. Walters, 1946.

8. BRUÏNE-ALTES, JOHANNA, C., DE, & HUIZING, H. C. The monaural masking method for recruitment testing in symmetrical deafness. *Acta Oto-laryng.*, Stockh., 1949, 37, 385–391.

9. BUSTAMENTE GURRIA, A., & GARIBAY, A Comentarios al fenomenon de Fowler. *Pr. med. Mex.*, 1944, 9, 124–125.

10. DAVIS, H., MORGAN, C. T., HAWKINS, J. E., JR., GALAMBOS, R., & SMITH, F. W. Temporary deafness following exposure to loud tones and noise. Report No. PBM 57250, Comm. on Med. Sciences, OSRD, Sept. 30, 1943. Available in microfilm or photostat from the Office of Technical Services, Dept. of Commerce, Washington 25, D. C. Summarized in *Laryngoscope*, St. Louis, 1946, 56, 19–21 and in *Acta Oto-laryng.*, Stockh., 1950, Suppl. 88, 57.

11. DAVIS, H., STEVENS, S. S., NICHOLS, R. H., JR., HUDGINS, C. V., MARQUIS, R. J., PETERSON, G. E., & ROSS, D. A. *Hearing Aids.* Cambridge: Harvard Univer. Press, 1947.

12. DAY, K. M. Ménière's disease: present concepts of diagnosis and management. *Ann. Otol.*, etc., St. Louis, 1950, 59, 966–979.

13. DENES, P., & NAUNTON, R. F. Methods of audiometry in a modern deafness clinic. *J. Laryngol.*, 1949, 63, 251–275.

14. DENES, P., & NAUNTON, R. F. The clinical detection of auditory recruitment. *J. Laryngol.*, 1950, 64, 375–398.

15. DIX, M. R., HALLPIKE, C. S., & HOOD, J. D. Observations upon the loudness recruitment phenomenon, with especial reference to the differential diagnosis of disorders of the internal ear and VIII nerve. *Proc. roy. Soc. Med.*, 1948, 61, 516–526.

16. DIX, M. R., HALLPIKE, C. S., & HOOD, J. D. "Nerve" deafness: its clinical criteria, old and new. *Proc. roy. Soc. Med.*, 1949, 42, 527–536.

17. EBY, L. G., & WILLIAMS, H. L. Recruitment of loudness in the differential diagnosis of end-organ and nerve fiber deafness. *Laryngoscope*, St. Louis, 1951, 61, 400–414.

18. FLETCHER, H. Loudness, masking and their relation to the hearing process and the problem of noise measurement. *J. acoust. Soc. Amer.*, 1938, 9, 275–293.

19. FLETCHER, H., & MUNSON, W. A. Loudness, its definition, measurement, and calculation. *J. acoust. Soc. Amer.*, 1933, 5, 82–108.

20. FOGHT, E. Bemerkungen über das akustische Leiden bei der Ménière'schen Krankheit. *Acta Oto-laryng., Stockh.*, 1944, Suppl. **51**, 230–240.

21. FOWLER, E. P. Marked deafened areas in normal ears. *Arch. Otolaryngol., Chicago*, 1928, **8**, 151–155.

22. FOWLER, E. P. A method for the early detection of otosclerosis. *Arch. Otolaryngol., Chicago*, 1936, **24**, 731–741.

23. FOWLER, E. P. The diagnosis of diseases of the neural mechanism of hearing by the aid of sounds well above threshold. *Trans. Amer. otol. Soc.*, 1937, **27**, 207–220 (with discussion by Lorente de Nò). Also in *Laryngoscope, St. Louis*, 1937, **47**, 289–300 (reprinted without discussion).

24. FOWLER, E. P. Measuring the sensation of loudness. *Arch. Otolaryngol., Chicago*, 1937, **26**, 514.

25. FOWLER, E. P. The use of threshold and louder sounds in clinical diagnosis. *Laryngoscope*, 1938, **48**, 572–588.

26. FOWLER, E. P. Concerning certain hearing phenomena. *Ann. Otol., etc., St. Louis*, 1941, **50**, 576–578.

27. FOWLER, E. P. Hearing standards for acceptance, disability and discharge in the military services and industry. *Trans. Amer. otol. Soc.*, 1941, **31**, 243–263. Also in *Laryngoscope, St. Louis*, 1941, **51**, 937–956.

28. FOWLER, E. P. A simple method of measuring percentage of capacity for hearing speech. *Arch. Otolaryngol., Chicago*, 1942, **36**, 874–890.

29. FOWLER, E. P. A method for measuring the percentage of capacity for hearing speech. *J. acoust. Soc. Amer.*, 1942, **13**, 373–382.

30. FOWLER, E. P. The recruitment of loudness phenomenon. *Laryngoscope, St. Louis*, 1950, **60**, 680–695.

31. FOWLER, E. P., JR. (Ed.) *Medicine of the ear.* New York: Thomas Nelson, 1939.

32. GARDNER, M. B. Short duration auditory fatigue as a method of classifying hearing impairment. *J. acoust. Soc. Amer.*, 1947, **19**, 178–190.

33. GARNER, W. R. Hearing. In C. P. Stone (Ed.), *Annual review of psychology.* Vol. 3. Stanford: Annual Reviews, 1952. Pp. 85–104.

34. GEIGER, P. Loudness level to loudness conversion chart. *J. acoust. Soc. Amer.* 1940, **11**, 308–310.

35. GRADENIGO, G. Studie proposte de acumetria. Indice vocále (Index vocalis) Auditus. *Arch. ital. di Otolog.*, 1912, **23**, 177–208.

36. GUTTMAN, J., & HAM, L. B. Effect of an interfering tone upon the hearing of a normal and a deafened ear. *Phys. Rev.*, 1928, **31**, 712.

37. GUTTMAN, J., & HAM, L. B. Masking effects of an interfering tone upon a deafened ear. *Arch. Otolaryngol., Chicago*, 1930, **12**, 425.

38. GUTTMAN, J., & HAM, L. B. Masking effects of an interfering tone upon a deafened ear. *J. acoust. Soc. Amer.*, 1930, **2**, 83–93.

39. GUTTMAN, J., & HAM, L. B. Masking effects of an interfering tone upon a deafened ear. *Laryngoscope, St. Louis*, 1930, **40**, 648.

40. HABERMANN, J. Ueber die Schwerhörigkeit der Kesselschmiede. *Arch. Ohrenheilk.*, 1890, **30**, 1–25.

41. HALLPIKE, C. S., & HOOD, J. D. Some recent work on auditory adaptation and its relationship to the loudness recruitment phenomenon. *J. acoust. Soc. Amer.*, 1951, **23**, 270–274.

42. HALM, T. Determination of the difference limen and the latest illustration thereof in audiometry. *J. Laryngol.*, 1949, **63**, 464–466.

43. HARRIS, J. D. An historical and critical review of loudness recruitment. U. S. Navy Medical Research Laboratory Report No. 200, 1952, **11**. Pp. 47.

44. HARRIS, J. D., & MYERS, C. K. Loudness perception for pure tones and for speech. U. S. Navy Medical Research Laboratory Report No. 156, 1950, **9**, 97–127.

45. HARRIS, J. D., HAINES, H. L., & MYERS, C. K. Loudness perception for pure tones and for speech. *Arch. Otolaryngol., Chicago*, 1952, **55**, 107–133.

46. HARRIS, J. D., & RAWNSLEY, A. I. The locus of short-duration fatigue, or "adaptation." U. S. Naval Medical Research Laboratory Report No. 219, 1952, **11**.

47. HARRISON, M. S. Ménière's disease. *St. Thomas Hosp. Gaz.*, 1945, **43**, 88–99.

48. HAYES, C. S. Phonemic regression in relation to difference limens for pitch in the perceptively deafened ear. Unpublished Ph.D. thesis, Northwestern Univer., 1951.

49. HIDDEMA, A. Over de Index Vocalis. Thesis, Groningen University, Netherlands, 1928. (Not seen, cited in [7].)

50. HIRSH, I. *The measurement of hearing.* New York: McGraw-Hill, 1952.

51. HOLMGREN, L. Hearing tests and hearing aids. A clinical and experimental study. *Acta Oto-laryngol., Stockh.,* 1939, Suppl. 34.

52. HOOD, J. D. Studies in auditory fatigue and adaptation. *Acta Oto-laryngol., Stockh.,* 1950, Suppl. 92.

53. HOOD, J. D. Auditory adaptation and its relationship to clinical tests of auditory function. *Proc. roy. Soc. Med.,* 1950, 43, 1129–1136. Reprinted with discussion in *J. Laryngol.,* 1951, 65, 530–544.

54. HUGHSON, W., & WITTING, E. G. Estimations of improvement in hearing following therapy of deafness. *Ann. Otol., etc., St. Louis,* 1940, 49, 368–383.

55. HUIZING, H. C. Die Regressiefactor en Zijn Beteekenis Bij Het Functioneele Gehooronderzoeken de Gehoorprothese. *Ned. tijdschr. Geneesk.,* 1941, 85, 4529–4533.

56. HUIZING, H. C. Die Bestimmung der Regression bei der Gehörprufung und der physikalische, physiologische und psychologische Zusammenhang bei der Gehoorprothese. *Acta Oto-laryngol., Stockh.,* 1942, 30, 487–499.

57. HUIZING, H. C. Keuring van het Gehoororganaan. *Ned. Tijdschr. Geneesk.,* 1942, 86, 2580–2584.

58. HUIZING, H. C. The relation between auditory fatigue and recruitment. *Acta Oto-laryngol., Stockh.,* 1948, Suppl. 78.

59. HUIZING, H. C. The symptom of recruitment and speech intelligibility. *Acta Oto-laryngol., Stockh.,* 1949, 36, 346–355.

60. HUIZING, H. C. Perte auditive et intelligibilité. *Acta Oto-rhino-laryngol., Bruxelles,* 1949, 3, 367–375.

61. HUIZING, H. C., & BAKKER, R. De Bepaling der Regressie Bij Slechthoorendheid. *Ned. Tijdschr. Geneesk.,* 1942, 85, 2513–2516.

62. HUIZING, H. C., & POTHOVEN, W. J. Klassische und moderne Methoden bei der funktionellen Gehooruntersuchung. *Z. Hals.-Nas.-u. Ohrenheilk.,* 1941, 47, 390–401.

63. KATZ, F. G. Uber Sprachgehör und Sprachverstandnis. *Z. Hals-Nas.-u. Ohrenheilk.,* 1929, 25, 193–199.

64. KATZ, F. G., & SALIS, G. v. Quantitative Hörprüfung mit Sprache. *Z. Hals-Nas.-u. Ohrenheilk.,* 1930, 26, 106–126.

65. KOBRAK, F. New tests and clinical experiments on hearing. *J. Laryngol.,* 1940, 55, 405–423.

66. KOBRAK, F. Some clinical phenomena of deafness in the light of new and old tests of the ear. *Arch. Oto-laryngol., Chicago,* 1948, 47, 113–118.

67. KOBRAK, F. On the differential diagnosis, by Fowler's "Loudness Recruitment" test, of changes in the cochlear nerve, with reference to the "Loudness Shock Test." *Confinia Neurologica,* 1950, 10, 309–316.

68. LANGENBECK, B. Über die Unzulänglichkeit unserer Hörscharfenbestimmungen. *Z. Hals.-Nas.-u. Ohrenheilk.,* 1928, 20, 313–321.

69. LANGENBECK, B. Das Hören in Lärm. *Z. Hals-Nas.-u. Ohrenheilk.,* 1932, 30, 1–64.

70. LANGENBECK, B. Die Geräuschaudiometrie als diagnostische Methode. *Z. Laryng.,* 1950, 29, 103–121.

71. LANGENBECK, B. Die Geräuschaudiometrie als diagnostische Methode. II. Schwellennahe und schwellenferne Geräuschaudiogramme. *Z. Laryng.,* 1950, 29, 470–487.

72. LARSEN, B. Dansk Oto-Laryngologisk Selskabs Forhandlinger 1940–1941. (Not seen, cited in [7].)

73. LARSEN, B. *Nordisk Med.,* 1942, 14, 1714.

74. LASKIEWICZ, A. On the determination of intensity limen for recruitment phenomenon in bilateral deafness. *Acta Oto-laryngol., Stockh.,* 1950, 38, 403–416.

75. LIDÉN, G., & NILSSON, G. Differential audiometry. *Acta Oto-laryngol., Stockh.,* 1950, 38, 521–527.

76. LOWY, K. Functional studies on labyrinthine fenestration in animals. *Laryngoscope, St. Louis,* 1945, 55, 6–19.

77. LURIE, M. H. Studies of acquired and inherited deafness in animals. *J. acoust. Soc. Amer.,* 1940, 11, 420–426.

78. LÜSCHER, E. Zur funktionellen Einteilung der Schwerhorigkeit und zur funktionellen otologischen Diagnostik. *Pract. Oto-rhino-laryngol.,* 1948, 10, 254–266.

79. LÜSCHER, E. Die Unterschiedsschwelle für Tonintensitätsänderungen bei psychogenen Nörstörungen. (Zur funktionelltopischen Diagnostik des nervosen Apparates des Ohres.) *Pract. Oto-rhino-laryngol.,* 1949, 11, 107–113.

80. LÜSCHER, E. The difference limen of intensity variations of pure tones and its

diagnostic significance. *Proc. roy. Soc. Med.*, 1950, **43**, 1116–1128.

81. LÜSCHER, E. Nouvelle méthodes d'examen de l'audition. *Bull. Acad. Suisse de Sci. Med.*, 1950, **6**, Suppl. I, 177–187.

82. LÜSCHER, E., & ERMANNI, A. Die topisch diagnostische Bedeutung der Unterschiedsschwelle fur Tonintensitätsänderungen. *Arch. Ohr.-, Nas.-, Kehlkopfheilk*, 1950, **157**, 158–216.

83. LÜSCHER, E., & ZWISLOCKI, J. Über An- und Abklingvorgange. *Ohres. Experientia*, 1945, **1**, 1–3.

84. LÜSCHER, E., & ZWISLOCKI, J. Eine einfache Methode zur monauralen Bestimmung des Lautstärkeausgleiches. *Arch. Ohr.-, Nas.-, Kehlopfheilk.*, 1948, **155**, 323–334.

85. LÜSCHER, E., & ZWISLOCKI, J. Eine einfache Methode zur monauralen Bestimmung des Recruitment phenomenon (Lautstärkeausgleich). *Pract. Otorhino-laryngol.*, 1948, **10**, 521–522.

86. LÜSCHER, E., & ZWISLOCKI, J. A simple method for indirect monaural determination of the recruitment phenomenon. *Acta Oto-laryngol., Stockh.*, 1948, Suppl. 78, 156–168.

87. LÜSCHER, E., & ZWISLOCKI, J. Adaptation des Ohres an Schallreize als Mass für die Lautstärkeempfindung und die Erregungsverteilung im Cortischen organ. *Acta Oto-laryngol., Stockh.*, 1949, **37**, 498–508.

88. LÜSCHER, E., & ZWISLOCKI, J. Comparison of the various methods employed in the determination of the recruitment phenomenon. *J. Laryngol.*, 1951, **65**, 187–195.

89. DE MARÉ, G. Audiometrische Untersuchungen. *Acta Oto-laryngol., Stockh.*, 1939, Suppl. 31, 173.

90. DE MARÉ, G. Fresh observations as to the so-called masking effect of the ear and its possible diagnostic significance. *Acta Oto-laryngol., Stockh.*, 1940, **28**, 314–316.

91. DE MARÉ, G. Investigations into the function of the auditory apparatus in perception deafness. *Acta Oto-laryngol., Stockh.*, 1948, Suppl. 74, 107–115

92. DE MARÉ, G., & RÖSLER, G. Untersuchungen über den Verdeckungseffekt bei Leitungs-und Innerohrschwerhörigkeit. *Acta Oto-laryngol., Stockh.*, 1950, **38**, 179–190.

93. MEISTER, F. J. Physikalische Fragen und Gesichtspunkte bei der Diagnostischen Geräusch-Verdeckungsmessung am Ohr. *Akustische Beihefte*, 1952, Heft 2, 49–58.

94. MEYER ZUM GOTTSBERGE, A. Zur Physiologie der Haarzellen. *Arch. Ohr.-Nas. u. Kehlkopfheilk*, 1948, **155**, 308–314.

95. MYERS, C. K. Recruitment of speech in certain types of deafness. *J. Speech and Hearing Disorders*, 1949, **14**, 293.

96. MYGIND, S. H. Ein Versuch zur Erklärung des sogenannten Regressions Phänomens (recruitment). *Z. Laryngol. Rhinol.*, 1950, **29**, 277–285.

97. NEUBERGER, F. Untersuchungen über den qualitativen Zusammenhang der Unterschiedsschwelle für Tonintensitätsanderungen und dem Lautstärkeausgleich. *Mschr. Ohrenheilk.*, 1950, **84**, 169–182.

98. NILSSON, G. Some aspects of the differential diagnosis of obstructive and neural deafness. *Acta Oto-laryngol., Stockh.*, 1948, **30**, 125–138.

99. PIRODDA, E. Il compartamento della soglia di discriminazione per l'intensita' in conduzione ossea. *Oto-rino-laringol., Ital.*, 1950, **19**, 152–158.

100. POHLMANN, A. G., & KRANZ, F. W. Binaural minimum audition in a subject with ranges of deficient acuity. *Proc. Soc. exp. Biol., N. Y.*, 1924, **21**, 335–337.

101. POHLMAN, A. G., & VERMILLION, S. D. Unreal phenomena in audition and their relation to test methods. *Ann. Otol., etc., St. Louis*, 1933, **42**, 352–364.

102. POTHEVEN, W. G., & SCHURINGA, A. Aviation noise deafness, hearing standards and recruitment. *J. aviat. Med.*, 1940, **19**, 380–388.

103. REGER, S. N. Loudness level contours and intensity discrimination of ears with raised thresholds. *J. acoust. Soc. Amer.*, 1935, **7**, 73.

104. REGER, S. N. Differences in loudness response of the normal and hard-of-hearing ear at intensity levels slightly above the threshold. *Ann. Otol., etc., St. Louis*, 1936, **45**, 1029–2039.

105. SALTZMAN, M. *Clinical audiology*, New York: Grune and Stratton, 1949.

106. SALTZMAN, M. Recruitment phenomenon in craniocerebral trauma. *Arch. Otolaryngol., Chicago*, 1950, **52**, 782, 785.

107. SCHUBERT, K. Hörermüdung und Hördaner. *Z. Hals-, Nas.-u. Ohrenheilk.*, 1944, **51** 19–74.

108. SCHUBERT, K. Ein neues Audiometer und die damit erzielten Ergebnisse. *Z. Laryngol. Rhinol.*, 1951, **30**, 11–26.

109. SHUTTS, R. E. Differential sensitivity to frequency change in the perceptively deafened adult. Unpublished Ph.D. thesis, Northwestern Univer., 1950.

110. STEINBERG, J. C. The function of the apical turns of the cochlea and the symptoms of a lesion in this location. IV. Discussion from the point of view of the physicist. *Trans. Amer. otol. Soc.*, 1935, 25, 131–135.

111. STEINBERG, J. C. Nerve deafness of known pathology or etiology. In Symposium: the neural mechanism of hearing. V. Etiological and clinical types of so-called "nerve deafness." *Laryngoscope, St. Louis*, 1937, 47, 603–611.

112. STEINBERG, J. C., & GARDNER, M. B. The dependence of hearing impairment on sound intensity. *J. acoust. Soc. Amer.*, 1937, 9, 11–23.

113. STEINBERG, J. C., & GARDNER, M. B. On the auditory significance of the term hearing loss. *J. acoust. Soc. Amer.*, 1940, 11, 270–277.

114. STEVENS, S. S., & DAVIS, H. *Hearing.* New York: Wiley, 1930.

115. TONNDORF, J. Auditory perception in aircraft noise. USAF Sch. Aviat. Med., Project No. 21-27-001, Report No. 1, Dec. 1951.

116. TONNDORF, J. The relation of pure-tone threshold to speech perception in white noise. USAF Sch. Aviat. Med., Project No. 21-27-001, Report No. 2, Dec. 1951.

117. TUMARKIN, A. A. A contribution to the theory of the mechanism of the auditory apparatus. *J. Laryngol.*, 1945, 60, 337–368.

118. TUMARKIN, A. The decibel, the phon, and the sone. *J. Laryngol.*, 1950, 64, 179–188.

119. VECKMANS, J. L. Essai d'explication du "recruitment." *Acta oto-rhino-laryngol., Bruxelles*, 1949, 3, 355–360.

120. VECKMANS, J. L. Techniques modernes d'audiometrie. *Scalpel, Bruxelles*, 1950, 103, 1315–1319.

121. VEIS, J. Flustersprache und Konversationssprache in ihren Beziehungen Zueinander. *Arch. Ohrenheilk.*, 1913, 90, 200–211.

122. WATSON, L. A. *A Manual for advanced audiometry.* Minneapolis: Colorcraft Press, 1949.

123. WATSON, L. A., & TOLAN, T. *Hearing tests and hearing instruments.* Baltimore: Williams and Wilkins, 1949.

124. WATSON, N. A., & KNUDSEN, V. O. Selective amplification in hearing aids. *J. acoust. Soc. Amer.*, 1940, 11, 406–419.

125. WEERSMA, P. Dissertation, Groningen University, 1938. (Not seen, cited in [126].)

126. WEERSMA, P. Über den Einfluss von Lärm auf das Tongehör und Schlechthorender. *Z. Hals-Nas.-u. Ohrenheilk.*, 1941, 47, 402–411.

127. ZANGEMEISTER, H. E. Über den Lautstärkeausgleich. *Acta Oto-laryngol., Stockh.*, 1950, 38, 484–497.

128. ZWAARDEMAKER, H. *Handelingen XIV Nederl.* Natuur- en Geneeskundig Congres, Gehouden te Delft, Maart 1913. (Not seen, cited in [7].)

129. ZWAARDEMAKER, H. Ondersoekingen in het Physiol. Laboratorium te Utrecht, Ve reeks, No. 17, 1916. (Not seen, cited in [7].)

ON THE INTERACTION OF SIMULTANEOUS RESPONSES[1]

DONALD R. MEYER[2]

The Ohio State University

In this discussion a theory will be developed to account for the effects of induced muscular tension upon learned and unlearned responses. It will then be shown that the theory permits evaluation of the role of tension in skill learning, and provides a rationale for the use of the blink rate as an index of generalized muscular tension. Finally, the theory will be related to the results of two recent studies of eyeblink conditioning to stress the immediate systematic importance of response interaction.

As the terms shall be used here, a muscle is tense if antagonistic effectors prevent it from shortening, contracted if the mechanical component is present as well. Muscular tension is said to be generalized if widespread muscle groups are active, focalized if just a few are involved. Many different methods have been devised for measuring muscular tension (26, 60). Some yield only indices of magnitude, but others show locus and patterning

as well. We shall emphasize the results of those investigations which have been designed to provide the more precise information.

By definition, both tension and contraction are properties of muscle that depend upon innervation. Distinctions between the two are largely arbitrary, for muscular activity is basically homogeneous. Thus a pattern of muscular tension, like a pattern of muscular contraction, is a response. Responses interact if they are modified upon simultaneous elicitation. Such modifications pose the general problem with which we shall deal.

TWO MOLAR LAWS OF INTERACTION

To begin with, there are two major factors which determine degree of interaction in artificial induction situations. First, interaction varies directly with the magnitude of the inducing response. For example, the amplitude of the knee jerk varies considerably as a function of changes in the activity of the other parts of the body (43, 59, 63, 111). Courts (17) has quantified this phenomenon. When his subjects squeezed a hand dynamometer with increasing amounts of force, knee jerk amplitude systematically increased.

Experiments by Freeman suggest another law, that interaction varies directly with the proximity of the simultaneous responses. He measured rate of finger oscillation and varied the locus of induced muscular tension. Maximal facilitation of rate was obtained if the adjacent finger was the site of induction (33, 35).

In a further study finger oscillation was produced by stimulating motor points and again muscular tension was most effective if located in nearby muscle groups (36).

These facts can be related to information derived from the study of isolated responses. The degree to which a distant muscle group is involved in a response varies directly with the magnitude of the response and the proximity of the group to the maximally activated effectors. For example, Freeman (33) had subjects flex their fingers and recorded activity in other body parts. Changes were detected when the fingers were maximally flexed, but not when they were minimally flexed.

Hines (56) has pointed out that this increased distribution of muscular activity can be demonstrated by palpation. If fingers are flexed minimally, only muscles of the hand and lower arm are tense. If flexion is powerful, muscles as distant as those which extend the neck and adduct the scapula are involved. Shaw (87) has shown how extensive these changes can be by measuring the distribution of activity in gripping a hand dynamometer. He found changes in action potentials, many of them marked, in the opposite arm, hand, and leg, the foot on the same side, the mouth, jaw, nose, ear, chest, abdomen, and neck.

The proximity relation can be illustrated by results obtained by Davis (25), who measured electrical changes in the forearms and left leg of subjects while they worked mental multiplication problems or learned nonsense syllables. During multiplication subjects tend to move their writing hand, and Davis found that the changes were greatest here and least in the left leg. Smaller changes were recorded during nonsense sylla-

ble learning, presumably because the focus of activity has been moved to muscles remote from all loci of measurement.

These facts show that interacting responses are overlapping responses. The factors which determine the probability that each of two simultaneous responses will involve a given muscle group are the same as those which determine degree of interaction. Our problem, then, is to account for the way in which overlap is translated into interaction.

The most parsimonious assumption is that interaction depends upon the convergence of simultaneous patterns of neural impulses. This hypothesis is the cornerstone of the present theory. Fortunately, the locus of the convergence is not difficult to specify, for the molar laws of interaction suggest that the event takes place in the nuclei of the motor system.

The Plan of Representation within the Motor System

Relevant to this conclusion are neurophysiological experiments in which the input to the motor system is systematically varied and the output of the system is observed. Usually the input is an electrical shock delivered to a cortical motor area. The output is analyzed through study of the muscular activity which results.

If the magnitude of the input to the motor system is kept at threshold value, the characteristics of the output vary as a function of the locus of the input. Points can be found in the cerebral cortex from which isolated movements of a finger, elbow, or toe can be obtained. A recent analysis of the precentral motor cortex of the monkey by Woolsey and Settlage (119) shows that there is a remarka-

ble correspondence between place in the motor system and place in the musculature. In fact, a chart of the motor cortex which summarizes the responses elicited from different points resembles a somewhat distorted, but intact monkey.

Similar information has been obtained for human subjects during routine stimulation of the cerebral cortex in neurosurgery. The most recent chart of threshold response data has been prepared by Penfield and Rasmussen (76). It shows a plan of motor representation which can be summarized by a somewhat disjointed figure with a small trunk and small legs, but large hands and feet. Centers for vocalization, salivation, and mastication overlap the ones for lips, jaw, tongue, and swallowing.

A further feature of the organization of the motor system is revealed if the stimulus is fixed in locus and varied in magnitude. Suppose that a point has been identified which, when a threshold stimulus is applied, gives an isolated flexion of the finger. If stronger shocks are delivered to this point, progressively more remote muscle groups are brought into play. The movement first involves the fingers alone, then the fingers and wrist, and then the fingers, wrist, and elbow. Very strong shocks elicit widespread movements. If a different point is stimulated, a similar sequence results in which the activity of a given muscle group depends with its proximity to the threshold group.

These data show that there is a spatial gradient of representation for each body part. A given movement is most readily obtained if a particular point in the motor cortex is stimulated, but around this focus there is a fringe region from which the movement can be elicited in combination with others. The closer the fringe point is to the focus, the greater its sensitivity. Thus the motor channel for a body part varies in extent with the strength of the stimulus that is used in searching for the bounds.

It must be a rare input that fires only those cells which make up a representational focus. Distributed excitation within the motor system is the rule rather than the exception. The factors which are responsible for distribution deserve special consideration. Because of them, motor pathways are seldom discrete even when the input is sharply defined.

Input-output analysis yields none of the details of organization which are responsible for the overlapping somatotopic plan. Chang, Ruch, and Ward (14) suggest that one mechanism is cortical, that each muscle of the body is represented by a field of efferent neurons which varies in density from the focus to the fringe. When a focus is activated with a threshold stimulus, none of the fringe cells for neighboring muscles are fired. When the stimulus applied to the focus is more intense, excitation spreads and progressively more cells in the surrounding fields go into action.

Fibers within the cortex itself seem to be involved. In a study of the facilitation of a response to stimulation of one point by stimulation of a different point, Graham Brown (48) found that interaction disappeared if knife cuts were made between the two electrode positions. Foerster (32) reported that irradiation to the fingers could be prevented if a stimulated thumb area were completely surrounded by a cut. Finally, Erickson (30) found that epileptic seizures can be altered in course if a trench divides upper and lower halves of the sensorimotor cortex.

Distribution of excitation also takes place at lower levels. One line of evidence comes from examination of conditions under which a response to stimulation of one cortical point will recur when a different point is subsequently activated (72). Since cutting intrahemispheric connections has no effect upon such facilitation, it must be due to persistence in subcortical structures.

The fact that reflex irradiation (89) can be demonstrated in the bulbospinal preparation also indicates subcortical distribution. There is new evidence, too, which suggests that corticospinal tracts participate, for Glees and Cole (45) report that restricted lesions in the hand area of the monkey motor cortex result in degeneration of the tracts all the way down the spinal cord.

Feedback excitation from proprioceptive end organs within muscles is both a contributor and distributor of excitation within the motor system. Gellhorn (42) recorded muscle action potentials during stimulation of the motor cortex and found that magnitude of the electromyogram was greatly increased if the observed muscle was fixed in a lengthened position. Gellhorn (43) also showed that proprioceptive feedback from one muscle can increase the reactivity of neighboring muscles.

This feedback excitation undoubtedly affects several levels, and cerebral cortical areas are definitely involved. Woolsey, Chang, and Bard (116) stimulated the afferent spinal nerve roots of the monkey and found that the electrical activity is evoked in the precentral as well as the postcentral areas of the brain. This somatic afferent projection is larger than the one defined by light tactile stimulation (118), for under the latter conditions the potential changes are limited to the postcentral regions.

The new potentials related to deep sensitivity can be distinguished from the potentials evoked in the postcentral region by a light tactile input. In the precentral gyrus they are so distributed that Woolsey, Chang, and Bard suggest that proprioceptive projection fibers terminate in the motor area for the same body part. But the effects of proprioceptive feedback are not restricted to the focus for the muscle which contains the receptor. It is a characteristic property of afferent systems that localized peripheral stimulation gives rise to activity within a considerable area of the cortex, with maximal changes in a central focal region (84, 105, 118, 121).

Thus the stage is set for proprioceptive feedback to alter the properties of two strategically important components of the motor system. Thus far the existence of motor areas other than the one in the precentral gyrus has not been considered, but actually there are at least four which supply somatic musculature (40, 41, 47, 76, 117, 120). The functions of the different systems have not yet been specified satisfactorily, but this is not essential for the present discussion. Although the characteristics of movements elicited from each area vary considerably, the overlapping somatotopic plan of organization has been found for all that have been studied in detail.

No attempt is made to evaluate the relative importances of these complementary distributive mechanisms. It should also be pointed out that the processes discussed are not the only ones which contribute to lack of definition within a motor pathway or to variations in width of a channel. The input to the motor system almost certainly is not dis-

crete unless it is artificial, and it probably varies in extent with changes in the magnitude of the output.

LIMITATIONS UPON THE INTEGRATIVE PROPERTIES OF THE MOTOR SYSTEM

The forms of movement which can be observed if the motor system is activated by an artificial stimulus have not yet been discussed, but the generalization is reasonably straight-forward. If the input to the motor system is simple, the output of the motor system is simple. If a punctate, threshold electrical shock is delivered to an area for the wrist, we can observe flexion, extension, or deviation. The complexity of the movement can be altered by a change in the intensity of the stimulus, but the resulting activity of wrist, fingers, and arm has a limited resemblance to the movement involved in the throwing of a baseball. In short, a skilled movement is never observed, which suggests that the motor system has properties which cannot be altered by practice.

It has been proposed that the fibers which form a meshwork about the efferent cells in the sensorimotor cortex might be modified during acquisition, but Sperry (97) has shown that intersecting knife cuts in this region alter motor capacities of monkeys very little. Evidently the fibers contribute little to the patterning of a response. There is a note-worthy parallel between Sperry's findings and the surprisingly small changes in the spatio-temporal organization of human responses after therapeutic section of the corpus callosum (10, 94, 95, 96, 113). Perhaps a distributive role can be assigned to both classes of fibers. There are facts of bilateral interaction which add further support to the notion.

The Sperry data, unless one postulates an unacceptable independence of integration and neural connections (62), indicate that the pattern of the output of the motor system is largely dependent upon the pattern of the input. This hypothesis is hardly new; it is similar in conception to Tower's (107) keyboard analogy for the pyramidal tract. It is supported by clinical data and by experiments like those of Loucks (64) and Harlow and Bromer (52).

In summary, it is assumed that the motor system is not involved in the modifications of the central nervous system which take place during the acquisition of a skill. The motor system is a transducer which makes definite contributions to the synthesis of movements, but the properties of the system are fixed. Cerebellar loops are included within the system as defined here, and from the periodicity of motor behavior (53, 77, 100, 115), it is concluded that they operate as an input-informed feedback mechanism (83).

INDUCTIVE FACILITATION OF UNLEARNED RESPONSES

Consider next the interpretation of the Courts (17) experiment in terms of the principles just discussed. When the patellar tendon is tapped a motor channel of definite extent is activated, and thresholds of neurons in the vicinity of this channel are altered. But the input does not capture extensor neurons alone because of the overlap within the motor system and by the lack of sharp definition within the input itself. Hence the occurrence of extension is a statistical proposition; the input activates a population of cells in which there is greater representation of extensors than of flexors.

When the dynamometer is

squeezed another channel is activated, and its extent is proportional to the force of the grip. Excitation is distributed to the vicinity of the leg channel, and trips off some of the neurons that are near, but not at threshold when hand activity is absent. It has no effect upon the cells that are triggered by the input for the knee jerk.

The differential recruitment of extensor neurons is a function of the changes brought about by the patellar input, for the contribution of distributed excitation from the hand is relatively minor and cannot be expected to have a selective effect. Now, it is apparent from the spatial gradient of representation that the fringe regions adjacent to the leg channel are predominantly extensor in function. Similarly, the closer the cells are to the pathway, the more they are changed by the patellar input. Hence distributed excitation from the hand channel fires more extensor than flexor neurons, and the result is an increase in jerk amplitude. Facilitation increases up to the point where as many flexor as extensor neurons are recruited. An inversion of the function takes place if and when distributed excitation fires the relatively remote pools of neurons that are dominated by flexor cells.

With an assumption that reaction criteria are met when a critical population is brought to threshold, the same concepts account for the fact that induced tension alters the latency of a response (35, 38), or the rate and magnitude of a repetitive response. For example, finger tremor rate increases under load, and French (39) reports that finger tremor amplitude goes up when subjects squeeze dynamometers with the opposite hand. Body sway is more pronounced if one stands with the muscles of the

legs or body generally tensed (92, 93). No attempt is made to account for the periodicity of these responses; the treatment is limited to alterations of responses that are already established.

In summary, induced tension facilitates the responses just discussed by making the motor system available to an input, and not by operating upon the input. The theory is an obvious extension of concepts introduced by Sherrington (90) in his classical treatment of spatial summation. It can account for changes in the magnitude or the latency of a response. Presumably there are no other kinds of inductive facilitation.

INDUCTIVE FACILITATION OF LEARNED RESPONSES

If all that muscular tension can do is alter the magnitude or latency of a response, it is obvious that one must forego the notion that it can have a direct effect upon the formation of habits. Several experiments which support the opposite view require reinterpretation. The pioneer study is that of Bills (2), who found that nonsense syllable performance improved if subjects squeezed dynamometers. Although Bills' trends were not marked, subsequent studies have amply confirmed them.

But Stauffacher (98) and Courts (18) have shown that the interaction is not always facilitatory. They varied the degree of induced tension and found that there is first an increase and then a decrease in facilitation as grip pressure increases. Under conditions of extreme induction, Courts obtained interference.

According to Freeman (34), the results mean that proprioceptive feedback from induced muscular tension lowers the thresholds of excitability for all levels of functional

activity in the nervous system. The increased excitation makes the whole brain work better up to a point where the system is overloaded. A similar viewpoint is sponsored by Bills (3).

Actually, however, there is little evidence for such widespread effects. There can be no doubt that proprioceptive input alters the level and distribution of excitation within the motor system, but here the fiber bundles are concentrated into a projection system. So far as correlational mechanisms are concerned, proprioceptive feedback is only one kind of input among many. Thus it is probable that any unusual properties of the proprioceptive input are a function of the peculiar position of the proprioceptive projection.

In a discussion of several alternative interpretations of the Bills experiment, Robinson (81) held that the results might be attributed to an increased readiness to react in all muscle groups. This theory is similar in many respects to one just proposed to account for the facilitation of simple responses. Courts (20) has noted that research has substantiated the spread of muscular tension to effectors other than those directly involved in the process of induction, but adds that no one has yet been able to show how such generalized tension operates to produce both facilitation and interference.

However, by holding that interference is produced by facilitation of irrelevant responses, it is possible to account for the phenomena observed in complex performance situations with the same theory that handles facilitation of the knee jerk. To borrow the terminology of Hull (58), the difference observed by Bills is one of reaction potential and not of habit strength.

If this hypothesis is correct, then interactive facilitation cannot be demonstrated unless there is some differentiation of the input in favor of the criterion response. Once selection has taken place, a channel is activated with a fringe that has predictable characteristics. From the spatial gradient of representation, the neurons whose thresholds have been altered most would contribute to the response, for they are situated adjacent to the focus of activity. Hence small amounts of added excitation associated with hand activity are most likely to fire these neurons, and facilitation is the result. If more excitation is added, progressively more remote pools of neurons are recruited. Eventually performance declines because a given amount of induction recruits more inappropriate than appropriate neurons.

Obviously, the effects of induced tension cannot be the same throughout practice. If an input captures an optimal population, added excitation triggers only neurons which contribute to the production of errors. This conclusion is documented by the data of Courts (21). He studied changes in the functional relation between degree of tension and performance as learning takes place. The results for nonsense syllable learning and pursuit rotor learning were similar. The greatest amounts of facilitation were obtained early. The optimal degree of tension became progressively less and the detrimental effects of extreme tension became more pronounced as practice proceeded. Freeman (37) reports a similar finding in mirror drawing.

Evidently, then, muscular tension can be a valuable asset during the early stages of practice on a verbal or motor task, but a hindrance to skilled performance. Thus Russell (85) found that muscular tension

interferes with ball-tossing performance, particularly in men. For the skilled individual any instruction that enhances tenseness produces errors. In motor skills, error production usually appears as an increase in response variability.

Several studies illustrate this relation. Bousfield (9) found that tremor rate, amplitude, and variability all increase during ergographic work. Renshaw and Postle (79) report that pursuit tracking is impaired if the stylus is gripped tightly. Furthermore, Feldman (31) found that two measures of variability in a tracking task tended to increase as a function of the degree of induced tension.

It is difficult to reconcile these facts with results which Bills (2) reports for the task of adding columns of digits. He found that hand tension improves both speed and accuracy. The facilitation of latency is understandable, but only if accuracy remains the same or deteriorates. There is little consensus on this point, for Freeman (35) reports greater speed with less accuracy in continuous addition and Zartman and Cason (122) report no effect.

The data of Block (7) are of relevance here. She used hand pressure, foot pressure, and the two combined for the induction conditions; arithmetic problems, syllogisms, and analogies were the materials. Facilitation could not be demonstrated. This supports the assumption that tension cannot facilitate selective processes. Bills and Stauffacher (4) provide further evidence. Their subjects supported weights and solved arithmetic reasoning problems or analyzed short detective stories; the only differences were in latency of response.

In summary, consider an observation which, if duplicated with proper controls, is catastrophic for the present argument. Bills (2) presented lists of paired associates for a single trial under conditions of normal and induced tension, and then studied recall with no induced tension. The group which learned under tension recalled the list slightly, but significantly better than the normal control group. However, the list required only thirty seconds for presentation and the recall trial followed almost immediately. It is doubtful that relaxation was achieved and that induced changes were completely dissipated (29).

This experiment illustrates another necessary control if the facilitated response is learned. A response can operate upon the input for another only if its occurrence alters the stimulus situation to which the simultaneous response is conditioned. In the Bills experiment this factor should favor the group which both learned and recalled without tension, so the phenomena cannot be explained in this way. Under other circumstances a change in magnitude or latency might be attributed either to a change in the stimulus situation or to a loss of interactive facilitation.

It should be possible to design adequate experiments to isolate these factors. One approach to the problem can be derived from the assumption that strength of conditioning, as distinguished from response strength (55), should not vary as a function of changes in proprioceptive feedback. The latter argument follows from the results of experiments by Grant and Schneider (49, 50) on the effects of changes in the magnitude of conditioned stimuli.

SIGNIFICANCE OF CHANGES IN THE DISTRIBUTION OF MUSCULAR TENSION DURING LEARNING

The foregoing theoretical analysis can be extended to learning situations in which artificial tension is absent. Under these circumstances the criterion response can be visualized as interacting with all the others, even though the latter may be difficult to define separately. However, these extraneous responses are unlike an induced response in that they are eliminated as a function of practice.

The sequence of elimination is such that tension changes are from generalization to focalization (28). Sometimes this process can be traced not only to a single limb, but within the limb itself as well. For example, Renshaw and Schwarzbek (80) found that in hand pursuit learning the trunk muscles first relax, and then those of the shoulder, upper arm, and lower arm in order. Distant muscle groups are frequently involved in the changes. Freeman (33) noted a progressive reduction in the tension of a leg muscle as subjects learned a pursuit task or nonsense syllables.

Changes in focal muscle groups are complicated. The input for the response increases with practice, but distributed excitation derived from irrelevant responses decreases. These facts preclude simple interpretation of studies in which measures of tension are derived from the active limb. In this group are the investigations of Daniel (22), Telford and Swenson (103), Stroud (99), and Ghiselli (44).

Recall that facilitation by induction is most pronounced early in learning, that the optimal amount of tension is progressively less as learning proceeds, and that late-in-learning tension interferes with performance. The ideal conditions, then, are such that tension is maximal early in

learning and declines progressively as a function of practice. Note that these conditions prevail when there is no artificial tension, even though the amount of tension at a given point in the practice series might be below the optimal level. Early in learning the musculature is widely activated, and the responses eliminated first interact the least with the criterion response. Elimination results in loss of facilitation, but at the same time the occurrence of the criterion response is less dependent upon it.

Muscular tension is typically generalized under most of the conditions which we commonly associate with difficult performances. This relation was first pointed out by Duffy (28), and its use in the analysis of work situations has been explored by Ryan, Cottrell, and Bitterman (86). Factors thus far identified are lack of practice, prolonged work periods (1, 82), distractions (23, 73), complex materials (24, 88), and increased incentives (15, 35). According to the present theory, in each of the instances cited additional tension contributes to response elicitation.

It is a common practice for flying teachers and other motor skill instructors to stress the value of relaxation to their students. Yet it is evident that a positive, as well as a negative, role can be assigned to the tenseness of individuals who have just begun to learn. To say that tenseness should not be tampered with may seem foolish to an individual who has long associated the effortless with the expert performance, but the available evidence indicates that optimal conditions for the novice are not the same as those for the skilled performer. A datum directly related to this conclusion has been reported by Davis (25), who found that instructions to relax the right

arm interfered with nonsense syllable learning.

THE BLINK RATE AS AN INDEX OF GENERALIZED MUSCULAR TENSION

In most of the experiments just cited, elaborate instrumentation was employed to detect changes in the distribution of muscular tension as learning proceeds. However, several lines of evidence support the conclusion that the blink rate can be used as an index of degree of generalization. Telford and Storlie (102), for example, had subjects learn a mirror-drawing task. They found that there is a reduction in the blink rate as a function of practice.

This observation can be correlated with the fact (76) that the locus of the motor channel for the eyelid is in a particularly strategic position. It is bordered on one side by the massive structures for the tongue and face; on the other by the huge representation for the hand. Hence a man makes few responses "without batting an eye." In animals, Blount (8) notes that blink rate increases during certain infrequent facial movements, e.g., yawning, sneezing, and eating. Thus, the blink response is facilitated by simultaneous activity in neighboring motor channels, and most important motor channels fall in this category.

Several other phenomena support this interpretation. Difficulty is a factor which alters distribution of tension, and Clites (16) found that blink rate varies with the difficulty of a nonvisual task. Recitation of the alphabet brings about an increase in rate, but the effect is more pronounced if the recitation is backward. Other data show a positive relation between the blink rate and another distributive factor, the incentive for performance (15).

In the latter experiment, blink rate measurements were taken in the interval between trials in a tracking task. The particular conditions are emphasized because the frequency of blinks during performance was almost zero. The task required continuous visual control, a condition which inhibits the blink rate. Drew (27), for example, studied the blink frequency of subjects who tracked target or drove an automobile. He observed that the blink rate decreases as a subject moves from heavy traffic onto the open road, or tracks a target which follows a course of increased complexity.

Luckiesh and his collaborators (68) have studied the effects of visual working conditions upon the blink rate. Their task, reading, did not require continuous scrutiny. Illumination, type size, presence or absence of glare, and the visual correction worn were variables related to the rate of blinking. Significantly, Luckiesh and Moss (66, 67) have also found that changes in muscular tension accompany changes in glare and illumination. The relationship of blink rate to illumination during reading is in question, however, for MacFarland, Holway, and Hurvich (69), Tinker (106), and MacPherson (71) have failed to confirm the trends.

Bitterman (5) studied the effects of type size and with Soloway (6), the effects of auditory distractions on the blink rate of subjects given a variant of the Minnesota Test for Clerical Workers. They were unable to demonstrate any significant changes. Bitterman and Soloway pointed out that auditory distractions have been shown to be accompanied by increases in muscular tension, and concluded that muscular tension does not affect blinking.

But, as Luckiesh (65) points out,

direct comparisons between his results and those of Bitterman cannot be readily made. In Bitterman's visual task, pairs of identically or dissimilarly spelled words are presented to be judged alike or different. This material lacks the redundancy of English prose, for each letter must be examined if performance is to be efficient. Even in reading there is a reduction in the blink rate (104), and this change takes place in spite of the fact that blinks tend to occur at periods and page turns (51).

Other kinds of visual work are not so conveniently parcelled. MacPherson (71) found that the blink rate, as compared to a control condition, went down most for counting rows of dots or for detecting aircraft silhouettes, somewhat less during reading, and least during the plotting of graphs. MacPherson concluded that the blink rate is inversely correlated with the ease of the task. But, since Clites (16) found the opposite relationship in a nonvisual task situation, the effects of difficulty are probably obscured by the requirement of continuous visual input.

Since fatigue is one of the conditions in which muscular tension is distributed, studies of blink rate in relation to prolonged visual work are relevant. As one might expect in a situation where antagonistic processes operate, the results are variable. Hoffman (57), Carpenter (13), and Tinker (106) report increasing blink rates; Carmichael and Dearborn (12), Simonson and Brozek (91), and Brozek, Simonson, and Keys (11) find no impressive trends. Wood and Bitterman (114) point out a fact which accounts for most of the differences: the blink rate does not change if the subject knows that this performance is being measured.

Only the Carpenter data cannot be handled by this hypothesis. They are understandable, however, if related to observations made with the Tufts alertness indicator (110). This instrument has been designed to provide automatic signals for human beings who might go to sleep under dangerous circumstances. It measures muscular action potentials from just above the eye. The choice of site was an empirical one, for it was found that the activity of this muscle reflected the level of activity in other effectors. Why this should be so is evident from the present theory.

Travis and Kennedy (108) showed that as action potentials from the eye decline, hand reaction times to visual and auditory stimuli go up. Similarly, Kennedy and Travis (61) demonstrated that action potentials and tracking performance are correlated. The same investigators (109) found that the index varies with the task, and significantly for the problem at hand, that it is high during simulated lookout performance that requires a hand reaction.

In the Carpenter study, the subjects were tested with the Mackworth apparatus (70), which requires a key-pressing response each time a pointer on a dial moves a double step instead of a single step. Thus Carpenter's situation is one in which more excitation is distributed to the eyelid mechanism than under reading conditions and the marked effects observed are to be expected.

Many factors must be considered by anyone who attempts to use the blink rate as an index of generalized muscular tension. Nevertheless, the index has many advantages. There are savings in cost and time required for securing and analyzing records, and no equipment need be attached to the subject. Furthermore, the subject need not be aware that meas-

urements are being taken, and so the blink rate index is particularly adaptable for use in clinical experimental situations.

FACILITATION OF THE ISOLATED BLINK

The present theory predicts that induced muscular tension should have an effect upon both the blink rate and the blink reflex. There are no data of this kind so far as rate is concerned, and the status of the reflex problem is far from clear. Freeman (37) reported that the eyelid reflex is facilitated by tension induced in the lower limbs, and Peak (74) obtained facilitation of blinks to loud sounds by induced hand activity. Courts (19), however, found no facilitation by hand tension if blinks were elicited by a puff of air. Peak (75) suggested that Courts' responses were maximal to begin with, that his subjects were not fully relaxed, or that time intervals are important. Tye (112) performed experiments to control these factors, and used both sounds and puffs of air for eliciting the blink. He interpreted his data in support of Courts.

However, it is probable that such facilitation was a factor in a recent experiment by Taylor (101). She administered to a population of college students an inventory of "manifest anxiety," and then studied eyeblink conditioning in two groups of subjects chosen from the tails of the distribution. She found that the frequency of conditioned blinks was greater in the anxious than in the nonanxious group.

One criterion of anxiety is an unusual amount of muscular activity. The supposition is evident in the choice of items for the inventory used by Taylor. This, together with the fact that the eyeblink representation is strategically located, suggests that the Taylor effect is a change in reaction potential attributable to response interaction. Taylor reports in a footnote a fact that is absolutely, if not statistically, gratifying: of the four subjects discarded because their unconditioned blinks were too frequent, three were classified as anxious.

The supposition that has been advanced is strongly supported by data obtained by Ponder and Kennedy (78). They studied the blink rates of subjects who were excited or angry, and found that the interval between blinks was sharply reduced. But another observation of these investigators is even more relevant. They went to a court of law and counted the blinks of individuals who were giving testimony. These people blinked much more frequently during cross-examination—a situation which would make almost every one of us anxious.

One final argument can be offered in support of the notion that Taylor's data represent an instance of interactive facilitation. It follows from the present theory that conditioned discrimination should be more difficult to develop in anxious subjects, for the facilitation should prolong the elicitation of a response that is being extinguished. This deduction has just been confirmed by Hilgard, Jones, and Kaplan (54). They used Taylor's inventory to obtain anxiety scores for a group of subjects. Subsequently an eyelid response was conditioned to light, and then a conditioned discrimination was developed between two different positions of the conditioned stimulus. There was a significant positive correlation between anxiety scores and lack of discrimination.

Summary

In review, the theory that has just been developed attributes the interaction of simultaneous responses to a convergence of impulse patterns upon the motor pathways of the central nervous system. It is nothing more than a combination of classical concepts of spatial summation and modern information on details of organization within the motor system.

The principal conclusion to be drawn from this analysis is that interactive facilitation can alter the magnitude or latency of a response, but has no direct effect upon response selection. One corollary of this notion is that induced muscular tension affects rate of acquisition by promoting response elicitation, and not by altering correlative mechanisms.

Interference with performance is held to result from a facilitation of competing responses. A basis for differential recruitment of motor neurons is found in the spatial gradient of motor representation for each body part. Changes in the effects of induced muscular tension from facilitation to interference are assumed to be related to changes in the probability that increments in excitation will facilitate more appropriate than inappropriate motor neurons.

It has been pointed out that the change in distribution of muscular tension during learning is invariably one of focalization. From this it is concluded that the degree of natural tenseness varies in same way during learning as does the optimal degree of induced muscular tension. Since a positive, as well as negative, role can be assigned to this tenseness, instructors in motor skills should abandon the practice of stressing the value of relaxation.

A rationale has been developed for the use of the blink rate as an index of generalized muscular tension. Variations in the rate are attributed to excitation distributed to the motor channel for the eyelid from neighboring channels for the head and hand. Finally, it is suggested that similar facilitation accounts for the results of two recent studies of eyelid conditioning in anxious subjects.

For psychologists, the present argument is a molecular theory. We do not assume that molecular concepts are better than any other kind, but they seem to yield the most parsimonious interpretation of the data under discussion. We think that it has been shown that molecular theory can predict molar behavior and, furthermore, that molecular theory can indicate the minimum features of any molar theory. For example, in the Hull system there are many constructs, but one seems to be missing and required. If the present theory is correct, it could best be labelled *efferent neural interaction*.

REFERENCES

1. Ash, I. E. Fatigue and its effect upon control. *Arch. Psychol.*, 1914, No. 31.
2. Bills, A. G. The influence of muscular tension on the efficiency of mental work. *Amer. J. Psychol.*, 1927, **38**, 227–251.
3. Bills, A. G. *The psychology of efficiency.* New York: Harper, 1943.
4. Bills, A. G., & Stauffacher, J. C. The influence of voluntarily induced tension on rational problem solving. *J. Psychol.*, 1937, 4, 261–271.
5. Bitterman, M. E. Heart rate and frequency of blinking as indices of visual efficiency. *J. exp. Psychol.*, 1945, 35, 279–292.
6. Bitterman, M. E., & Soloway, E. The relation between frequency of

blinking and effort expended in mental work. *J. exp. Psychol.*, 1946, **36**, 134–136.

7. BLOCK, H. Influence of muscular exertion upon mental performance. *Arch. Psychol., N. Y.*, 1936, No. 202.

8. BLOUNT, W. P. Studies of the movement of the eyelids of animals: blinking. *Quart. J. exp. Physiol.*, 1928, **18**, 111–125.

9. BOUSFIELD, W. A. The influence of fatigue on tremor. *J. exp. Psychol.*, 1932, **15**, 104–107.

10. BRIDGMAN, C. S., & SMITH, K. U. Bilateral neural integration in visual perception after section of the corpus callosum. *J. comp. Neurol.*, 1945, **83**, 57–68.

11. BROZEK, J., SIMONSON, E., & KEYS, A. Changes in performance and in ocular functions resulting from strenuous visual inspection. *Amer. J. Psychol.*, 1950, **63**, 51–66.

12. CARMICHAEL, L., & DEARBORN, W. F. *Reading and visual fatigue.* Boston: Houghton Mifflin, 1947.

13. CARPENTER, A. The rate of blinking during prolonged visual search. *J. exp. Psychol.*, 1948, **33**, 587–591.

14. CHANG, H. T., RUCH, T. C., & WARD, A. A., JR. Topographical representation of muscles in the motor cortex of monkeys. *J. Neurophysiol.*, 1947, **10**, 39–56.

15. CHRISTNER, C. Blink rate as an indicator of muscular tension. Unpublished master's thesis, The Ohio State University, 1951.

16. CLITES, M. S. Certain somatic activities in relation to successful and unsuccessful problem solving. *J. exp. Psychol.*, 1935, **18**, 708–724.

17. COURTS, F. A. The knee-jerk as a measure of muscular tension. *J. exp. Psychol.*, 1939, **24**, 520–529.

18. COURTS, F. A. Relations between experimentally induced muscular tension and memorization. *J. exp. Psychol.*, 1939, **25**, 235–256.

19. COURTS, F. A. The influence of muscular tension on the eyelid reflex. *J. exp. Psychol.*, 1940, **27**, 678–689.

20. COURTS, F. A. Relations between muscular tension and performance. *Psychol. Bull.*, 1942, **39**, 347–367.

21. COURTS, F. A. The influence of practice on the dynamogenetic effect of muscular tension. *J. exp. Psychol.*, 1942, **40**, 504–511.

22. DANIEL, R. S. The distribution of mus-

cular action potentials during maze learning. *J. exp. Psychol.*, 1939, **24**, 621–629.

23. DAVIS, R. C. The relation of certain muscle action potentials to "mental work." *Ind. Univer. Publ. Sci. Ser.*, 1937, No. 5.

24. DAVIS, R. C. The relation of muscle action potentials to difficulty and frustration. *J. exp. Psychol.*, 1938, **23**, 141–158.

25. DAVIS, R. C. Patterns of muscular tension during "mental work" and their constancy. *J. exp. Psychol.*, 1939, **24**, 451–465.

26. DAVIS, R. C. Methods of measuring muscular tension. *Psychol. Bull.*, 1942, **39**, 329–346.

27. DREW, G. C. Variations in reflex blink-rate during visual-motor tasks. *Quart. J. exp. Psychol.*, 1951, **3**, 73–87.

28. DUFFY, ELIZABETH. The relation between muscular tension and quality of performance. *Amer. J. Psychol.*, 1932, **44**, 535–546.

29. DUSSER DE BARENNE, J. B., & MCCULLOCH, W. S. Factors for facilitation and extinction in the central nervous system. *J. Neurophysiol.*, 1939, **2**, 319–355.

30. ERICKSON, T. C. Spread of the epileptic discharge. An experimental study of the after-discharge induced by electrical stimulation of the cerebral cortex. *Arch. Neurol. Psychiat., Chicago*, 1940, **43**, 429–452.

31. FELDMAN, A. C. The influence of experimentally induced muscular tension on variability of response. Unpublished master's thesis, University of Missouri, 1944.

32. FOERSTER, O. *Bumke und Foersters Handbuch der Neurologie.* Band 6. Berlin: J. Springer, 1936.

33. FREEMAN, G. L. The spread of neuromuscular activity during mental work. *J. gen. Psychol.*, 1931, **5**, 479–494.

34. FREEMAN, G. L. Mental activity and the muscular processes. *Psychol. Rev.*, 1931, **38**, 428–447.

35. FREEMAN, G. L. The facilitative and inhibitory effects of muscular tension upon performance. *Amer. J. Psychol.*, 1933, **45**, 17–52.

36. FREEMAN, G. L. The optimal locus of "anticipatory tensions" in muscular work. *J. exp. Psychol.*, 1937, **21**, 554–564.

37. FREEMAN, G. L. The optimal muscular tensions for various performances.

Amer. J. Psychol., 1938, **51**, 146–150.

38. FREEMAN, G. L., & KENDALL, W. E. The effect upon reaction time of muscular tension induced at various preparatory intervals. *J. exp. Psychol.*, 1940, **27**, 136–148.

39. FRENCH, J. W. A comparison of finger tremor with the galvanic skin reflex and pulse. *J. exp. Psychol.*, 1944, **34**, 494–505.

40. FRENCH, J. D., SUGAR, O., & CHUSID, J. G. Cortico-cortical connections of the superior bank of the Sylvian fissure in the monkey (*Macaca mulatta*). *J. Neurophysiol.*, 1948, **11**, 185–192.

41. GAROL, H. W. The "motor" cortex of the cat. *J. Neuropathol. exp. Neurol.*, 1942, **1**, 139–145.

42. GELLHORN, E. The influence of alteration in posture of limbs on cortically induced movements. *Brain*, 1948, **71**, 26–33.

43. GELLHORN, E. Proprioception and the motor cortex. *Brain*, 1949, **72**, 35–62.

44. GHISELLI, E. Changes in neuro-muscular tension accompanying the performance of a learning problem involving constant choice time. *J. exp. Psychol.*, 1936, **19**, 91–98.

45. GLEES, P., & COLE, J. Recovery of skilled motor functions after small repeated lesions of motor cortex in macaque. *J. Neurophysiol.*, 1950, **13**, 137–148.

46. GOLLA, F. L., & ANTONOVITCH, S. The relation of muscular tonus and the patellar reflex to mental work. *J. ment. Sci.*, 1929, **75**, 234–241.

47. GRAHAM BROWN, T. Motor activation of the post-central gyrus. *J. Physiol.*, 1914, **48**, 30–31.

48. GRAHAM BROWN, T. Studies in physiology of the nervous system. *Quart. J. exp. Physiol.*, 1915, **9**, 81, 101, 117.

49. GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38**, 690–696.

50. GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: II. The conditioned galvanic skin response to an auditory stimulus. *J. exp. Psychol.*, 1949, **39**, 35–40.

51. HALL, A. The origin and purposes of blinking. *Brit. J. Opthal.*, 1945, **29**, 445–467.

52. HARLOW, H. F., & BROMER, J. A. Acquisition of new responses during in-

activation of the motor, premotor, and somesthetic cortex in the monkey. *J. gen. Psychol.*, 1942, **26**, 299–313.

53. HICK, W. E. The discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, 1948, **1**, 36–51.

54. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.

55. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.

56. HINES, M. Significance of the precentral motor cortex. In P. C. Bucy (Ed.), *The precentral motor cortex.* Urbana: Univer. of Illinois Press, 1949. Pp. 459–500.

57. HOFFMAN, A. C. Eye movements during prolonged reading. *J. exp. Psychol.*, 1946, **36**, 95–118.

58. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.

59. JACOBSON, E. *Progressive relaxation.* Chicago: Univer. of Chicago Press, 1938.

60. JACOBSON, E. Muscular tension and the estimation of effort. *Amer. J. Psychol.*, 1951, **64**, 112–116.

61. KENNEDY, J. L., & TRAVIS, R. C. Prediction and control of alertness. II. Continuous tracking. *J. comp. physiol. Psychol.*, 1948, **41**, 203–210.

62. KÖHLER, W., & WALLACH, H. Figural after-effects: an investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.

63. LOMBARD, W. R. The variations in the normal knee-jerk. *Amer. J. Psychol.*, 1887, **1**, 5–71.

64. LOUCKS, R. B. The experimental delimitation of neural structures essential for learning: the attempt to condition striped muscle responses with faradization of the sigmoid gyri. *J. Psychol.*, 1936, **1**, 5–44.

65. LUCKIESH, M. Comments on criteria of ease of reading. *J. exp. Psychol.*, 1946, **36**, 180–181.

66. LUCKIESH, M., & MOSS, F. K. Muscular tension resulting from glare. *J. gen. Psychol.*, 1933, **8**, 455–460.

67. LUCKIESH, M., and MOSS, F. K. A correlation between illumination intensity and nervous muscular tension resulting from visual effort. *J. exp. Psychol.*, 1933, **16**, 540–555.

68. LUCKIESH, M., & MOSS, F. K. *Reading as a visual task.* New York: Van Nostrand, 1942.

69. MacFarland, R. A., Holway, A. H., & Hurvich, L. M. *Studies of visual fatigue.* Boston: Graduate School of Business Administration, Harvard Univer., 1942.

70. Mackworth, N. H. The breakdown of vigilance during prolonged visual search. *Quart. J. exp. Psychol.*, 1948, 1, 6–21.

71. MacPherson, S. J. The effectiveness of lighting—its numerical assessment by methods based on blinking rates. *Illum. Engng*, 1943, 38, 520–522.

72. McCulloch, W. S. Cortico-cortical connections. In P. C. Bucy (Ed.), *The precentral motor cortex.* Urbana: Univer. of Illinois Press, 1949. Pp. 211–242.

73. Morgan, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol.*, 1916, No. 35.

74. Peak, Helen. Intensity of stimulation and the amplitude and latency of the lid reflex. *Amer. J. Psychol.*, 1932, 44, 785–788.

75. Peak, Helen. Dr. Courts on the influence of muscular tension on the lid reflex. *J. exp. Psychol.*, 1942, 30, 515.

76. Penfield, W., & Rasmussen, T. *The cerebral cortex of man.* New York: Macmillan, 1950.

77. Peters, W., & Wenborne, A. A. The time pattern of voluntary movements. *Brit. J. Psychol.*, 1936, 26, 388–406; 27, 60–73.

78. Ponder, E., & Kennedy, W. P. On the act of blinking. *Quart. J. exp. Physiol.*, 1927, 18, 89–110.

79. Renshaw, S., & Postle, Dorothy K. Pursuit learning under three types of instruction. *J. gen. Psychol.*, 1928, 1, 360–367.

80. Renshaw, S., & Schwarzbek, W. C. The dependence of the form of the pursuit-meter learning function on the length of the inter-practice rests: I. Experimental. *J. gen. Psychol.*, 1938, 18, 3–16.

81. Robinson, E. S. Work of the integrated organism. In C. Murchison (Ed.), *A handbook of general experimental psychology.* Worcester: Clark Univer. Press, 1934. Pp. 571–650.

82. Robinson, E. S., & Bills, A. G. Two factors in the work decrement. *J. exp. Psychol.*, 1926, 9, 415–443.

83. Ruch, T. C. Motor systems. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 154–208.

84. Ruch, T. C. Sensory mechanisms. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 121–153.

85. Russell, J. T. Relative efficiency of relaxation and tension in performing an act of skill. *J. gen. Psychol.*, 1932, 6, 330–343.

86. Ryan, T. A., Cottrell, C. L., & Bitterman, M. E. Muscular tension as an index of effort: the effect of glare and other disturbances in visual work. *Amer. J. Psychol.*, 1951, 63, 317–341.

87. Shaw, W. A. The distribution of muscular action potentials during imagining. *Psychol. Rec.*, 1938, 2, 195–216.

88. Shaw, W. A., & Kline, L. H. A study of muscle action potentials during the attempted solution by children of problems of increasing difficulty. *J. exp. Psychol.*, 1947, 37, 146–158.

89. Sherrington, C. S. *The integrative action of the nervous system.* New Haven: Yale Univer. Press, 1906.

90. Sherrington, C. S. Remarks on some aspects of reflex inhibition. *Proc. roy. Soc.*, 1925, 97B, 519–545.

91. Simonson, E., & Brozek, J. Effects of illumination level on visual performance and fatigue. *J. opt. Soc. Amer.*, 1948, 38, 384–397.

92. Skaggs, E. B. Further studies of bodily sway. *Amer. J. Psychol.*, 1937, 49, 105–108.

93. Skaggs, E. B., Skaggs, I. S., & Jardon, M. Attention and bodily sway. *Amer. J. Psychol.*, 1932, 44, 749–755.

94. Smith, K. U. The role of the commissural systems of the cerebral cortex in the determination of handedness, eyedness, and footedness in man. *J. gen. Psychol.*, 1945, 32, 39–79.

95. Smith, K. U. Bilateral integrative action of the cerebral cortex in man in verbal association and sensori-motor coordination. *J. exp. Psychol.*, 1947, 37, 367–376.

96. Smith, K. U. The functions of the intercortical neurones in sensorimotor coordination and thinking in man. *Science*, 1947, 105, 234–235.

97. Sperry, R. W. Cerebral regulation of motor coordination in monkeys following multiple transection of sensorimotor cortex. *J. Neurophysiol.*, 1947, 10, 275–294.

98. Stauffacher, J. C. The effect of induced muscular tension upon various phases of the learning process. *J. exp. Psychol.*, 1937, 21, 26–46.

99. Stroud, J. B. The role of muscular ten-

sions in stylus maze learning. *J. exp. Psychol.*, 1931, **14**, 606–631.

100. TAYLOR, F. V., & BIRMINGHAM, H. P. Studies of tracking behavior. II. The acceleration pattern of quick manual corrective responses. *J. exp. Psychol.*, 1948, **38**, 783–795.

101. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

102. TELFORD, C. W., & STORLIE, A. The relation of respiration and reflex winking rates to muscular tension. *J. exp. Psychol.*, 1946, **36**, 512–517.

103. TELFORD, C. W., & SWENSON, W. J. Changes in muscular tension during learning. *J. exp. Psychol.*, 1942, **30**, 236–246.

104. TELFORD, C. W., & THOMPSON, N. Some factors influencing voluntary and reflex eyelid responses. *J. exp. Psychol.*, 1933, **16**, 524–539.

105. THOMPSON, J. M., WOOLSEY, C. N., & TALBOT, S. A. Visual areas I and II of cerebral cortex of rabbit. *J. Neurophysiol.*, 1950, **13**, 277–288.

106. TINKER, M. A. Involuntary blink rate and illumination intensity in visual work. *J. exp. Psychol.*, 1949, **39**, 558–560.

107. TOWER, S. S. The pyramidal tract. In P. C. Bucy (Ed.), *The precentral motor cortex.* Urbana: Univer. of Illinois Press, 1949. Pp. 149–172.

108. TRAVIS, R. C., & KENNEDY, J. L. Prediction and automatic control of alertness. I. Control of lookout alertness. *J. comp. physiol. Psychol.*, 1947, **40**, 457–462.

109. TRAVIS, R. C., & KENNEDY, J. L. Prediction and control of alertness. III. Calibration of the alertness indicator and further results. *J. comp. physiol. Psychol.*, 1949, **42**, 45–57.

110. TUFTS COLLEGE. Applications of electrophysiological techniques to human performance. The reading assessor—the alertness indicator. Tech. Report SDC 58-2-11, SDC Human Engineering Project 20-C-2 & 3, Contract N5ori-58 T. O. II, Project Designation NR-782-003, Tufts College, Institute for Applied Experimental Psych., 1950, **37**.

111. TUTTLE, W. W. The effect of attention or mental activity on the patellar tendon reflex. *J. exp. Psychol.*, 1924, **7**, 401–419.

112. TYE, V. M. The effects of experimentally induced muscular tension on the eyelid reflex. Unpublished master's thesis, Univer. of Missouri, 1945.

113. VAN WAGENEN, W. P., & HERREN, R. Y. Surgical division of the commissural pathways of the corpus callosum. *Arch. Neurol. Psychiat.*, 1940, **44**, 740–759.

114. WOOD, C. L., & BITTERMAN, M. E. Blinking as a measure of effort in visual work. *Amer. J. Psychol.*, 1950, **63**, 584–588.

115. WOODWORTH, R. S. The accuracy of voluntary movements. *Psychol. Monogr.*, 1899, **3**, No. 3 (Whole No. 13).

116. WOOLSEY, C. N., CHANG, H. T., & BARD, P. Distribution of cortical potentials evoked by electrical stimulation of dorsal roots in *Macaca mulatta. Fed. Proc. Amer. Soc. exp. Biol.*, 1947, **6**, 230. (Abstract)

117. WOOLSEY, C. N., & CHANG, H. T. Activation of the cerebral cortex by antidromic volleys in the pyramidal tract. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, **27**, 146–161.

118. WOOLSEY, C. N., MARSHALL, W. H., & BARD, P. Representation of cutaneous tactile sensibility in the cerebral cortex of the monkey as indicated by evoked potentials. *Johns Hopkins Hosp. Bull.*, 1942, **70**, 399–441.

119. WOOLSEY, C. N., & SETTLAGE, P. H. Pattern of localization in the precentral motor cortex of *Macaca mulatta. Fed. Proc. Soc. exp. Biol.*, 1950, **9**, 140. (Abstract)

120. WOOLSEY, C. N., SETTLAGE, P. H., MEYER, D. R., SENCER, W., PINTO, T., & TRAVIS, A. M. Precentral and supplementary motor representation and its implications for the concept of a premotor area. *Res. Publ. Ass. nerv. ment. Dis.*, 1952, **30**, 238–264.

121. WOOLSEY, C. N., & WALZL, E. M. Topical projection of nerve fibers from local regions of the cochlea to the cerebral cortex of the cat. *Johns Hopkins Hosp. Bull.*, 1942, **71**, 315–344.

122. ZARTMAN, E. N., & CASON, H. The influence of an increase in muscular tension on mental efficiency. *J. exp. Psychol.*, 1934, **17**, 671–679.

BECK, S. J. *Rorschach's test.* Vol. III. *Advances in interpretation.* New York: Grune & Stratton, 1952. Pp. viii+301. $5.50.

Dr. Beck has high standing among those who study the Rorschach in America, because of his steady interest in improving the technique and his desire to communicate clearly his procedures and theories. His new book presents his method of interpretation more completely than before. Those who are using the test or training others in it will find great use for the clear and intelligent discussions of theory and of four case analyses. All four of these persons were under treatment, but represent patients from the community rather than hospital patients. For two cases Beck has repeated tests, and for all cases has clinical notes made during treatment.

The "advances" in Beck's statements do not represent a marked change in viewpoint. As always, he insists upon regarding the examination as an opportunity to sample behavior and thought production, not as a device for generating scores. He makes this clear by commenting that Tredgold, with the Mare and Foal Test, could make the same type of dynamic interpretation. This point of view might seem to cut the ground from under most of the talk about interpretation of Rorschach per se; as would also the recent report that in interpreting the adjustment shown in protocols, a qualified psychologist who knew nothing about Rorschach could agree with trained Rorschachers as well as they could agree with themselves.[1] Beck, however, considers the Rorschach performance an unusually good source of revealing cues, and this book is an attempt to make those cues explicit for other testers.

The cues now employed differ little from those discussed in his earlier books. He elaborates more clearly than before and distinguishes, for instance, between the significance of a cue in patients and the interpretation of the same cue in superior nonpatients. Beck has explored many new leads, such as the Levy-Zubin analysis of movement. He points out nuances in each type of response which should qualify the interpretation of the main scores. He does not reverse any earlier interpretations.

This in itself should suggest that the book is disappointing. We have now had thousands of research studies, some well conducted, which have failed to establish validity of many interpretations commonly made. One would expect such evidence to be used in revising the interpretative scheme, or that positive evidence would be advanced to demonstrate the validity of Beck's statements as to the equivalence of test behavior and personality structure. Beck does suggest that this might be a later step, but many readers would prefer to find a psychological proposition accompanied by, or preceded by, the reason for affirming it.

Beck discusses in a few pages current research on the test. He criticizes some studies for trying scoring innovations instead of validating Rorschach's Rorschach test. These in-

[1] GRANT, Q., IVES, V., & RANZONI, J. H. Reliability and validity of judges' ratings of adjustment on the Rorschach. *Psychol. Monogr.*, 1952, 66, No. 2 (Whole No. 334).

novations have, I think, rarely been sufficient to obscure the significance of positive or negative findings.

His second criticism is more substantial, that attempts to validate single signs or scores do not take into account the interactions between cues that an interpreter might use. It is apparent, however, that Beck in this book makes many statements about the significance of single scores, i.e., about main effects. While these interpretations would be qualified by added facts from other scores or qualitative features, such additions are embellishments to a main interpretation that he implies would be true more often than false in persons generally.

We find Beck turning his back on research evidence in the way that is too common in the literature and in the conversation of clinicians, by saying, in effect, "Evidence or no evidence, these propositions have clinical validity" (p. 43). In view of Beck's desire to make his *operations* public so that the Rorschach test will be a part of psychology rather than a cult, it is regrettable that he does not make his *validation* public. He does not see that so-called clinical validity is a type of private, or subjective, validity and hence not acceptable save as a source of questions to be tested by research.

The matter of color shock will serve to document this criticism. The literature now contains several studies which make it doubtful that color shock is truly color shock, or that it has the correlates claimed for it. Beck however treats color shock in the traditional manner, going further to regard color shock as evidence of anxiety regarding present temptation, while shading expresses guilt regarding past misdeeds. A rather attractive rationalization for this position is offered. Beck then dismisses the negative findings of objective research on color shock by citing Siipola and others as if their work negated the other studies and supported his position.[2] An examination of the original will disclose that their findings do not substantiate conventional procedures for measuring color shock.

Before the Rorschach can hold an established position among those who are deficient in the "wish to believe," the interpretative statements such as Beck presents must be accompanied by explicit evidence. Such evidence would not say (to quote at random), "This inability to react to color is the mark of the person insensitive to the world's exhilarating values" (p. 45). It would say, "Persons who show no responses scored *C*, *CF*, or *FC* are judged to be emotionally unresponsive by such-and-such criteria *in x cases out of one hundred.*" Here is the crucial problem. Beck and others word statements as if generally true, although the context makes it clear that the interpretations are not true in 100 cases out of 100. Hence the propositions are not to be accepted as stated. This approach to the Rorschach does not leave us with testable claims. The Rorschach enthusiast can say that the truth is contained only in his many qualifications. But this too easily becomes an evasion of responsibility. If one tries to take all the qualifications into account, he arrives at a unique pattern of scores and cues for each record. One cannot find a reasonable number of cases with this pattern, and therefore cannot make the required assessment of the statement's *degree* of correspond-

[2] SIIPOLA, E., KUHNS, F., & TAYLOR, V. Measurement of the individual's reactions to color in ink blots. *J. Pers.*, 1950, 19, 153–171.

ence to reality. To verify a proposition we must know its probability of being true. As the propositions become more complex, they become unverifiable. If the main effects are indeed assumed in interpretation, it is *their* degree of validity which must be reported.

So much stress on what Beck's book lacks reflects my sense of proportion, not Beck's. He gives almost no space to these issues, compared to the extensive discussion of cases. Those discussions can be evaluated, perhaps, by some other clinician who would report whether they agree with his private experience. There are few persons who can match Beck in experience, or in the care with which he traces his own conscious thoughts. A reader might choose to test the statements in terms of plausibility. On this ground, Beck stands high. He gives a long train of argument to support each interpretation, rather than issuing it as from an oracle. He is painstakingly self-critical in specifics, while not being so critical of his major premises.

Judged from the viewpoint of interest and the quality of Beck's thinking about personality, the book is admirable. The book is frequently poetic. It is written with far more style than most of its contemporaries, and many passages are quotable. In criticizing Lewin's disinclination to the genetic approach, for instance: "Lewin's deficit for this kind of construct was that he was not clinically trained. He can think in two dimensions. But his anxiety as experimentalist blocks his moving into the third" (p. 11). Throughout, Beck does a good job of expressing in hundreds of illustrations the concept of personality as an interaction of many forces and perceptions within the person. Whether the Rorschach

is a dependable tool or not, a book like this affords an excellent basis for thinking about the extreme complexity of personality. It is a powerful antidote for an oversimple view of personality as described by a few common traits or factors. The book will perform this service, even for many Rorschachers, by debunking naive interpretation of the psychogram or the single numerical score.

In summary, the book appears to be a fine one for Rorschach users or students who plan to continue with the test. Beck has rich experience to report, and the fact that it is subjective does not prevent its being useful to other interpreters. The book will not change the opinion of those who presently regard most claims for the Rorschach as rationalizations or statements of faith. The book will be of little assistance to the minority who regard the validity of the test as an open question which must be settled by determining which interpretations hold up in what proportion of cases. Beck has an extensive research program in process, and it may be that his reports from them will belatedly give us reason to have confidence in the present book.

LEE J. CRONBACH

University of Illinois

SHOSTROM, EVERETT L., & BRAMMER, LAWRENCE M. *The dynamics of the counseling process.* New York: McGraw-Hill, 1952. Pp. xvi+213. $3.50.

This book is largely a how-to-do-it manual of counseling procedures in the setting of a student personnel program. Labeled *self-adjustive counseling*, the approach is a compounding of nondirective and directive viewpoints flavored with a dash of learning theory. The goal of

counseling is increasing self-direction on the part of the client. Highlighted is a structured sequence of client-centered interviews preceded by a group-orientation session and culminating in a synthesis interview. Records, tests, and vocational information are used when desired by the student.

Although the approach is claimed to encompass educational, vocational, and personal problems of adjustment, the reader will find that major attention is given to what might best be described as nondirective educational-vocational guidance. Neither psychotherapy nor projective tests is mentioned. "If the counselor feels he is counseling a student with problems too involved for his competencies, turning the student and his records over to the psychologist would be the safest action" (p. 106).

Two chapters are devoted to discussion of a control-group study designed to test the effectiveness of the self-adjustive approach vs. the traditional approach. Ratings of terminal evaluative interviews with regard to client satisfaction and feelings of self-direction favor the self-adjustive approach. Details such as number of cases are not provided.

While this readable little book contains many helpful suggestions, references, and devices, it is obviously too nontechnical and practical to serve as a basic text in the graduate training of clinical psychologists. For a broader audience, the book should be very useful and have wide appeal.

LEONARD S. KOGAN
*Institute of Welfare Research,
New York*

KERR, MADELINE. *Personality and conflict in Jamaica.* Liverpool, Univer. of Liverpool Press, 1952. Pp. xiii+221. 15 s.

This is another contribution to the growing literature in the field of culture and personality. Kerr has written a most interesting account of the effect of divergent culture upon the formation of personality. On the one hand, the natives of Jamaica are expected to conform to the standards and ways of life set down by the English who have dominated the island for more than a century. On the other hand, the Jamaican natives have their own culture which is often at odds with that imposed by the British.

Kerr defines personality "as a constellation of roles which the individual learns to play under given conditions" (p. 196). The basic difficulty with the native arises because the roles which he is expected to play are defined variously from the two cultures to which he is exposed. Among the many difficulties which the Jamaican faces in this world are first of all the prejudices related to color and class.

There is essentially a three-level class system in the island. At the bottom are the great mass of peasants and farm and city workers, most of whom are definitely Negroid. At the second level is the middle class made up of business groups, some professionals, and individuals who have made a success of farming. At the top is the social elite composed of higher officials, plantation owners, and other people of material substance and traditionally high family standing. The chapter entitled "Class and Color" points out how the attitudes toward color play a part at all class levels but particularly with people who have some Negro blood but who are trying to move upward in the social strata. Apparently the individual tends to react to color in terms of any given social constellation. "One day a person will

be bitter and anti-white, yet when another constellation is involved he will be concerned with fears and dislikes only of people darker than himself. It is as if in some constellations he is the almost white man with white ideals, in others he is the dark man resenting white domination" (p. 96).

The divergent cultures and what they do to the individual are nicely brought out in the discussion of the educational program. The native children are exposed to a curriculum borrowed from Great Britain. As a result the content of the courses is "mainly concerned with issues and facts which cannot possibly have any relation" to the pupils' everyday life. For example, they learn the botany of plants that they have never seen, nor ever will see, and the history of parts of the world that have absolutely no meaning to them. On the other hand, their own local history and materials on their own habitat are completely neglected.

With reference to the development of the personality, Kerr provides a wealth of interesting and informative material. She points out that the spontaneous extroverted characteristic of early childhood gradually gives way to a rigid and introverted pattern found in adolescence and in adulthood. The wells of spontaneous creativity tend to be dried up, the recreational life of the children, particularly that in the school, is practically nil, and the adolescent and the adult tend to develop an apathetic and withdrawn type of behavior.

The close relations of mother and child are examined and the important role of the grandmother is described. The latter is familiar with the magical practices associated with childbirth, supervises the event, and carries out the magical rituals which are deep in their culture. Kerr believes that the

place of the grandmother is a more or less natural consequence of the mother-oriented family in Jamaica. The function of the father is often rather tenuous, especially in the peasant class. A high proportion of these couples are not legally married and the husband and father often plays a quite incidental role in child care and discipline.

As might be expected, there is a considerable fantasy life among these people but, with few exceptions, it is not directed into creative activity. Rather, it runs off into supersitition, magic, and a great concern for religious emotionalism. There are outbreaks into trancelike states; there is much interest in religious revivalism, and occasionally there is an outright outbreak of violence. Another interesting feature of the fantasy life of the native is the rich verbiage to which he is addicted. Apparently this is an aspect of Negro culture everywhere, although no one has ever adequately explained it.

Another possible factor in this situation which is mentioned but not explored by the author concerns the high incidence of malaria and hookworm among these people. We know that apathetic behavior and a low level of energy expenditure follow from the recurrence and persistence of these two diseases. It seems to the reviewer that we might have here an interesting interplay of constitutional factors and a conflicting culture. That is, if these people were living in a more or less unified culture, the effect of malaria and hookworm might be evident, but their effects on the personality might be different. This hypothesis could be tested by some cross-cultural comparisons.

Finally, a word may be said about the methodology of this study. The data were principally from two sources. First, Kerr interviewed a

large number of people of all classes and kept protocols of such interviews. In addition she gave a number of projective tests to children and to a few adults. The projective techniques used were the Rorschach, the Lowenfeld Mosaic test, drawings, and a projective test devised especially by the author for Negro children. In general, the results of the tests confirmed the findings from the interviews.

The author closes her comments on the relationship between field work and the test results in words which serve as a summary statement of this book. "In general the tensions caused by economic frustration, color prejudice, wrong methods in education, produce social conditions of insecurity, both economic and psychological, of doubt over role functions and of encouragement of magical beliefs. These conditions in turn produce a basic personality of an unintegrated type" (p. 193).

KIMBALL YOUNG
Northwestern University

PIAGET, JEAN. *Play, dreams and imitation in childhood.* (Trans. by C. Gattegno and F. M. Hodgson.) New York: Norton, 1951. Pp. ix+296. $5.00.

The current work by Piaget is another stimulating and provocative contribution to the literature on the development of children's thinking. In this well-translated volume, Piaget has as his basic goal an explanation of the evolution of "representative activity," which is "characterized by the fact that it goes beyond the present, extending the field of adaptation both in space and in time" (p. 273). Such an activity is essential in reflective thought as well as in operational thought.

Two theses are presented by Piaget in the book: (*a*) the transition from rudimentary, primitive, and situational assimilation of experience to the operational and reflective adaptation of experience can be studied by the analysis of imitative behavior and play activity of the child from very early months of the life; and (*b*) various forms of mental activity —imitation, symbolic activity, and cognitive representation—are interacting.

Underlying his analysis is the assumption that two fundamental processes operate in development: assimilation and accommodation. Assimilation refers to the idea that "no new external element ever gives rise to perceptive, motor or intelligent adaptation without being related to earlier activities" (footnote, p. 80). Accommodation refers to the adaptation to the demands of reality. By virtue of both these processes the individual evolves schemas, which are considered by Piaget as elementary structures in the psychological life of the individual. New experiences of the organism are brought into these schemas by the assimilation and the accommodation processes. Assimilation and accommodation gradually achieve equilibrium in the course of growth, since equilibrium is achieved at the completion of growth.

With such an over-all conception as a guide for analysis, which is dynamic in nature, Piaget proceeds to show how imitation, play, and representative activity evolve. Consistent with his previous writings, we find here an evolutionary presentation with rough age groupings given for the various stages of development. Incidentally, the same general ranges are given as in previous writing, i.e., seven seems again a crucial year for the beginnings of certain well-defined stages of growth and eleven to twelve the beginnings of adult-like activity.

In the first section of the book he deals with imitative behavior. He presents evidence from the observation of his own children as data. The protocols are presented in the text at varying length as illustrative of his various "facts" and theoretical points. He shows that imitation follows through six stages until "true" imitation occurs which is a consciousness of imitation as well as an imitation controlled by the intelligence as a whole.

Following this is a discussion of play, in which the same type of data is presented. A theoretical consideration is given, using the assimilation and accommodation principles as basic guides. Stages of play behavior are described, which proceed from a purely functional pleasure stage which is primarily assimilative in nature to the eventual development of organized and socialized games with definite rules. Details of this evolution are discussed. An analysis of symbolic play is presented, which Piaget considers as primarily egocentric in nature. He also discusses the nature of the symbolic behavior and claims this derives clearly from the child's own thought, an assimilative activity.

In a treatment of secondary symbolism as seen in dreams, play, and unconscious activity, Piaget presents a penetrating analysis of classical Freudian thinking, taking issue with such concepts as the censor mechanism, the mechanism of repression, and infantile memory. His argument, however, is constructive in that he integrates some of the Freudian thought of ego, repressions, and the unconscious with his own explanations. Jung's concept of the relative generality of symbols and of symbolism as thought is also included in this explanation. Finally, Piaget comes to a

treatment of representative activity — cognitive representations — and shows the steps in development from the sensory-motor schemas to conceptual schemas. A step-by-step evolution is attempted.

The observations notwithstanding, this book seems to be largely speculative. It is packed full of provocative ideas and challenges, and certainly Piaget has fearlessly explored an area of great import. Some general limitations are apparent. There is little reference to current American or British research relevant to the area of play and symbolism. For example, there is no reference to modern theorizing and research in ego psychology and the relevant contributions from play therapy. Second, the speculative nature of the book limits its translation into testable hypotheses for replication or in some instances even additional data gathering by other psychologists. Third, the evidence presented by Piaget leaves much to be desired in terms of rigor, objectivity, and at times detail. Finally, the style is heavy, repetitious and even esoteric, which taxes the reader's comprehension and interest. Nevertheless this book offers a contribution to a vital area, the development of thought, which is in desperate need of investigating and has been too long neglected by American psychologists. Piaget has certainly trodden with a heavy step where many of us have feared to venture.

IRVING SIGEL

Merrill-Palmer School, Detroit

CATTELL, RAYMOND B. *Factor analysis. An introduction and manual for the psychologist and social scientist.* New York: Harper, 1952. Pp. xiii + 462. $6.00.

In his preface Cattell states that this book has been written to meet

the needs of (1) the general student who wishes to learn what factor analysis is about, (2) the instructor who wants a textbook in factor analysis as part of a course in statistics, and (3) the research worker in this field. There are, to be sure, several excellent books on the subject of factor analysis but the author feels that they are too advanced mathematically or too specialized for general use.

The book is divided into three parts. Part I deals with the general theory of factor analysis and its application as a scientific method. Factorization is represented as valuable in a preliminary clearing of the ground and as useful also in providing a better understanding of the variables which emerge from an experiment. Ways in which factors may be obtained from a matrix of correlations and the purpose and necessity for rotation are presented with emphasis upon the meaning of factor loadings. The thesis is developed that factors may be thought of as source traits when simple structure has been established. In addition to the usual procedures in which factors are obtained from the intercorrelations of tests (called R technique) or from the correlation of persons (called Q technique), Cattell introduces two new methods, the O and P techniques. In the P technique factors are determined from the correlations of test scores achieved by the same person over a period of time. When the time periods instead of the tests are correlated we have the O technique. In this terminology of matrix algebra O is the transpose of P. Both of these methods should be useful in studying trends over a period of time.

Considerable attention is given to the need for oblique factors in Part II, and methods are provided for the extraction of orthogonal and oblique factors by the centroid method. Cattell does not overlook other factor methods, but he places the main emphasis upon the centroid as being the most flexible and generally useful. In this judgment the reviewer concurs. The final section of the book is concerned with various problems which arise in the factorization of measurements of ability as well as of personality.

Cattell has met his third objective, namely, that of preparing a handbook for research workers, better than his first two. The difficulty level of the book is not adapted to the non-mathematically trained student. It is doubtful whether the beginner can read profitably beyond the first two chapters, and most graduate students in psychology will understand very little of Part II unless it is preceded by a review of analytic geometry and matrix algebra. The problem of rotation as presented in Chapter 12 is unnecessarily discursive; and the student is not likely to get from Chapter 13 a clear idea of the real need for and value of oblique factors. While Cattell has clarified the distinction between reference axes and factors, his treatment is hampered by a confusing terminology. Rules laid down for carrying out a technique are not sufficiently well illustrated to make the rules readily applicable. As an example, the procedures to be followed in making a multiple group centroid analysis (pp. 178–184) are set down without illustration in 11 steps and supplementary notes.

This book is a real contribution to the literature of factor analysis and should be read by everyone seriously interested in research in this field. The final chapter on short-cut pro-

edures will be especially valuable
to the research worker. A feature of
the book is its sprightly style, occa-
sionally refreshed with a touch of
humor.

HENRY E. GARRETT
Columbia University

WALLS, GORDON L., & MATHEWS,
RAVENNA W. *New means of study-
ing color blindness and normal
foveal color vision.* University of
California Publications in Psy-
chology. Vol. 7, No. 1. Berkeley:
Univer. of California Press, 1952.
Pp. iv+172. $2.50.

In Walls and Mathews' lively, un-
orthodox, and iconoclastic mono-
graph on color blindness, Figure 1
shows a schematic set of chromatic
response curves for the three-com-
ponent theory of color vision. It is
offered, according to the legend,
"only as an aid to the reader who is
not an expert in the field." Let the
nonexpert, however, beware. The
introduction includes detailed dis-
cussions of several theoretical issues
basic to the experiments and inter-
pretations developed later in the
work. There is a modification of the
Young-Helmholtz three-component
theory, one which the authors have
labelled the "excess hypothesis"; an
interpretation of protanopia as a
reduction or "loss" system and
deuteranopia as a fusion or "col-
lapse" system, an eclectic idea pre-
sented some years ago by Pitt; a dis-
cussion of the genetics of the more
common types of color blindness; a
history and critical evaluation of the
entoptic phenomenon known as
"Maxwell's spot," which leads
the authors to reject the orthodox
view that it results from absorption
by macular pigment; and finally the
novel idea is presented that the
entoptic Maxwell pattern is due ex-

clusively to the nonuniform distri-
bution of redness, greenness, and
blueness receptors in the fovea.

The experimental portion of the
monograph describes a battery of
four "tests" applied in the analysis of
color vision and its defects: the filter
anomaloscope, the so-called "3-light
test" (a simplified means for making
gross luminosity measurements), neu-
tral point determinations, and the
use of a purple dichromic filter for
eliciting the Maxwell spot phe-
nomenon. The latter, which con-
stitutes an entoptic demonstration of
the receptor-type distribution pat-
tern (RDP), is reported to be char-
acteristically different in protanoids
as compared with normals, and
completely absent in deuteranoids.
The third and final portion of the
book summarizes the results of the
differential diagnoses provided by
these tests and offers detailed case
analyses which are heavily inter-
larded with genetic interpretation
and speculation.

At a time when conformity domi-
nates the social and intellectual
climate it is refreshing to find long
accepted views and assumptions vig-
orously challenged. Furthermore,
the authors' strictures about the
indiscriminate and improper use of
macular pigmentation as an explana-
tory concept are very much in order,
and their claim that the RDP and
neutral point determinations can
serve as the basis for making a clear-
cut diagnosis between protanopes
and deuteranopes should stimulate
further inquiry along these lines.

Unfortunately, the text sometimes
suffers from selection and misinter-
pretation to support the authors'
bias. In reviewing this work, more-
over, a number of questions arise, all
basic to an evaluation of the various
hypotheses and findings presented.

For example, why should we believe that "When a protanope (or a deuteranope either, for that matter) calls something blue, one can 'take his word' for it. He is experiencing blue" (p. 87) when we have been told on page 6 that "to take the protanope's word for it . . . that the long-wave hue he calls yellow is the normal's yellow experience, is like asking a congenitally deaf man whether Brahm's Fourth or Beethoven's Third is the greater symphony." If we accept the conclusion that "Nothing is found to conflict with the supposition that the basis of each of the hereditary types of color blindness is retinal and receptoral" (p. 156), are we to discount the subject who is diagnosed as "a victim of *central protanopia*" (p. 130, their italics)? If we accept as valid the conclusion that "Besides standard (loss) protanopia, and central protanopia (in an occasional heterozygote), *collapse protanopia* apparently exists" (p. 162, their italics), aren't we stretching beyond useful recognition the basic explanatory dichotomy which says protanopia = loss, deuteranopia = collapse? If we are told (pp. 31–32), with respect to the size of the rod-free area, that "The most recent and best values today are those of Rochon-Duvigneaud (1943)" (30' to 40'), are we to ignore the fact that Rochan-Duvigneaud's 1943 text refers only to measurements originally reported by him in 1907, and that Polyak's authoritative modern research, which is not even mentioned by Walls and Mathews, gives a conflicting value of 1°40'? When we read "No conceivable amount of macula pigment could possibly change the hue of a monochromatic light for anyone, be he tetartanopic, normal, or whatever" (p. 42), are we to deny the existence of the Bezold-Brücke hue shift with change in intensity? If Walls and Mathews seek to reconcile Ahlenstiel's conflicting data with their own by gratuitously assuming that nearly all of Ahlenstiel's 94 color defectives were protanoid (p. 89) and that his normal group must have included many deuteranomals (p. 36), should we grant these assumptions on pure faith? If, for an observer like Ba. L., who "can make out no RDP at all," we are offered the conclusion (p. 131) that "she does have an objective, retinal, receptor-type distribution pattern like her twin's, or more likely like their mothers' . . . " but that "she has no basis for perceiving it entoptically," what real diagnostic value does the entoptic receptor distribution pattern retain?

This jungle of fact and fancy is indeed stimulating but it is clearly for the specialist.

LEO M. HURVICH
Eastman Kodak Company

BERRIEN, F. K. *Practical psychology.* (Rev. Ed.) New York: Macmillan, 1952. Pp. xv+640. $5.00.

What is practical psychology? Impractical psychology? What does practical itself mean? Dr. Berrien answers the first question by implication in what he includes in the text. Answers to the latter two questions remain unanswered. In *Practical Psychology* the author covers selectively what others have called applied psychology, and what, still earlier, was subsumed under the label of psychotechnology. Dr. Berrien has organized his book around five major areas or topics: problems of adjustment, industrial psychology, applications to consumers and advertising, crime, and personal problems. The first and last

isted topics, however, might well have comprised a single part of the book. Certainly the two chapters covering personal problems, one on vocational guidance and the other on effective speaking and writing, would not be out of place if added to such chapters as study efficiency, mental health and guidance, and adjustments in later life—these latter being among those comprising the first part.

Turning to specifics, the good features of the text are many; the bad ones relatively few. Dr. Berrien has done a splendid job of collating and summarizing empirical findings on work efficiency and accidents, on psychological research in advertising, and on the general area of psychology applied to crime. All chapters abound with relevant illustrative material. The undergraduate student, the intended consumer, will derive great benefit from a careful reading and translation into practice of what he finds on "how to study" and "principles of mental health."

On the negative side, the most pervasive shortcoming is the failure to set a theoretical matrix for the discussion of each topic. For example, in the chapter on training in industry and business, virtually all the discussion deals with the organization and mechanics of training without much attention to the fundamental learning process for which the organization and mechanics are ancillary. Similarly, in treating morale, no conceptual framework (e.g., communality between individual and group values, group-shared goals, or Maier's frustration principle) is given within which to interpret the specific findings cited, say, with regard to what workers want to get out of their jobs. In short, the book is replete with "what" and "how," but relative-ly sparse on "why." Patently, content and methodology in our rapidly expanding field are more subject to obsolescence in the light of new findings than the underlying why of behavior.

The book represents a major revision of the pre-World War II edition. The changes comprise rewriting and expansion, rather than thorough coverage of the more recent literature. While current citations occur, they are far outnumbered by early references. The author has turned out a readable, teachable book, which should be useful at the undergraduate level if curriculum organization includes a survey-type course in applied psychology. It will be in competition with the similar works of Burtt, Hepner, Husband, and others.

W. J. E. CRISSY

Queens College

VINACKE, W. EDGAR. *The psychology of thinking.* New York: McGraw-Hill, 1952. Pp. vii + 392. $5.50.

Although psychologists have long been concerned with the psychology of thinking, we find in the literature little systematic presentation and evaluation of the over-all problems of thinking. Thus the event of a text in this area should be of great interest. Unfortunately, we have here little more than a survey of literature. The author at the outset points out that he "has not been guided by preference for any particular species of theory" (p. vii). This becomes only too apparent when in the body of the book we find him using constructs from opposing points of view and unsuccessfully trying to reconcile them. This makes of course for a most eclectic and, at the same time, confusing presentation, too typical of many of our text books.

A clear-cut definition of thinking

is never offered by the author. In the introductory chapter he offers Warren's definition with some analysis and criticism, but does not clearly commit himself regarding its acceptability. Throughout the book, there are various statements which might be considered definitions of thinking, and which presumably represent the author's view. First, "Thinking is the use or reorganization or application of what has been learned" (p. 42); next, it is " . . . manipulation of the environment wholly or partially without overt activity . . . there is always some symbolic activity" (p. 58); and also, "Thinking serves in the adjustment of the organism to its environment, both internal and external" (p. 6). The book's organization and the author's justification for it do allow us to gather something of Vinacke's concept of the psychology of thinking. He distinguishes first of all between reasoning and imagination as two kinds of thinking, the former being more related to the external world and more controlled, the latter less so. Second, all thinking is personalized and shaped by selective and regulatory systems which are established through learning. These systems include concepts, attitudes, sets, traits, and motives. Finally he accepts the field theory interpretation of mental activity " . . . as occurring in the midst of a complex pattern of nervous activity" (p. 5). The larger divisions of the book contain a section on Reasoning, another on Imagination, and a third on Personalizing Functions of Thought. The subdivisions include chapters on concept formation, logical thinking, transfer, problem-solving, autistic thinking, creative activity, and internalization of experience.

In the chapter on the mechanism of thinking the author subscribes to the motor theory of thinking, offering Watson's view as one acceptable to him with only slight modification. He says, "Thinking consists in the reactivation of past experience by means of the implicit activities of the muscles originally involved in that experience" (p. 69). As evidence for this theory he offers the works of Jacobson and Max and others which have reported that action potentials can be obtained in the parts of the body about which the subject is instructed to think. He does recognize the possible criticism that such evidence does not prove that the movements are essential to the thought but may only be accompanying phenomena. In rejecting this interpretation, he offers the evidence which Jacobson reported, on the basis of one case, that a man with an amputated limb could not imagine or think of activities in that limb. There is since the work of Jacobson however, considerable work on phantom limb phenomena which argues against the author's interpretation.

The eclecticism and resulting confusion are most profound in the chapters on problem-solving, transfer, and the internalization process. In discussing transfer and retroactive inhibition he offers Katona's research as a valuable contribution in the area of transfer, particularly because of Katona's emphasis on the role of understanding and meaning. Next he offers a number of behavioristic notions about transfer and retroactive inhibition, and then with apparent enthusiasm devotes a section to the gestalt concept of trace systems and the forgetting process. He then concludes that "There is, however, no real conflict between the more static interference theory, and the dynamic, field view. Rather they complement each other" (p. 150). A

footnote on the same page states that there is essential agreement between the psychoanalytic and field theories in the interpretation of the forgetting process.

Again, Vinacke states in the section on internalization of experience that conditioning forms the basis for all the early training of the infant. At the same time, he points out that it is difficult to explain "long-term" learning by conditioning, as it is rather specific and subject to extinction. Conditioning then sets the groundwork upon which other learning mechanisms, such as "symbolic functions," take over. How they take over or how they are learned is never too clear. Furthermore, a discussion earlier in the book of the need for recentering and re-organizing in problem-solving situations is never reconciled with the concept of conditioning. At the end of one chapter he does leave the reader with a provocative question in this connection: "To what extent is recentering in problem-solving linked with, or independent of, generalization gradients?" (p. 155). In general then, it can be said that if Vinacke has a point of view which actually does bring into harmony the concepts he uses, he has not stated it explicitly.

Despite all of the above criticisms, this book has much of positive value to offer. The objections to an inadequately developed and organized point of view should not take away from its potential value in the classroom. It is a very well written book, and does contain a well-organized and rather complete survey of the literature. Some chapters are handled especially well, such as "Creative Activity" and "Ideas, Imagery and Imageless Thought." The latter part of the book devoted to imagination and the personalized aspects of

thought appears to be the best.

<div style="text-align:right">IRENE R. PIERCE</div>

Wellesley College

AUSUBEL, DAVID P. *Ego development and the personality disorders.* New York: Grune & Stratton, 1952. Pp. xi+564. $10.00.

Ausubel's first purpose, presumably, is to construct a comprehensive nonpsychoanalytic theory of personality and of the personality disorders. To this reviewer, his theory is "comprehensive" as a listing of the current and standard criticisms of Freud, and "nonpsychoanalytic" in rejecting id and superego. That is, his original contribution to theory is a questionable modification of what Freud called "mental topography."

However, he does cover a wide selection of the criticisms directed at psychoanalytic theory. He reformulates, redefines, rephrases, and reannotates in the light of recent knowledge. His suggestions are often sound and may make the psychoanalytic model more meaningful to psychologists and more amenable to objective research. Also, his analysis of "identification"—a major feature of the book—is thought-provoking, especially in his specific developments.

In regard to topography (id-ego-superego), Freud said: "It must not be supposed that these very general ideas are presuppositions upon which the work of psychoanalysis depends." Apparently Freud saw his rough mapping of the "mental apparatus" only as a crude model or analogy, convenient for communicating certain ideas. Ausubel offers an alternate model, largely in terms of ego development. In other respects, his theory is at least "Freudomorphic." For example, he objects to Freud's use of the term "instinct" and to his

pansexual emphasis. However, his revisions in such details, while extensive and familiar, fail to effect any essential change in the psychoanalytic model.

In contrast to Freud, Ausubel would put less emphasis on id-ego conflict and, perhaps more important, less emphasis on conscious-unconscious conflict. In contrast to Mowrer, he would put less emphasis on ego-superego conflict. He reduces personality dynamics to intra-ego conflict. He would have us see conflict and psychodynamics largely in terms of difficulties arising in various stages of ego development. Tendencies which are appropriate to one stage of ego development must be abandoned or modified for wholesome development in later stages. If the abandonment or modification is not satisfactorily accomplished, the residual tendencies—no longer appropriate— lead to both internal and external difficulties in adjustment. The degree of consciousness or unconsciousness is less important than Freud supposed. In general, instead of id-ego-superego disagreements, he sees less precisely labelled stages of ego development (never reified!) bargaining, conflicting, and compromising.

In dealing with Freud's topography, Ausubel's approach may be illustrated by his treatment of the id. Instead of the id, he uses such terms as (1) visceral needs, (2) regal omnipotence, and (3) hedonistic urges. The strength of visceral needs is related to body type; Sheldon's viscerotonic endomorph may have stronger needs than the ectomorph. Presumably much of what the analysts see as id may be explained in terms of the impatient "regal omnipotence" of very early infancy, which develops more or less universally because the parent does everything for the newborn child. However, this unreasonable expectation of quick gratifications is best understood as an early aspect of ego development. Hedonistic urges cover in general the early and lasting tendency to short term goals and immediate gratifications. This tendency is partly axiomatic and partly a learned pattern of early stage ego development. The analysts' id concept contaminates this tendency with other, misleading implications (instinct, sexuality).

Ausubel's focus is the ego. He might have advanced theory further if he had abandoned ego as well as id and superego. To him the ego is not the person, it is an identifiable abstraction "referring to the interrelated set of experiences, perceptions, attitudes, motives and values which revolve about the awareness of self." If he had dealt with the individual functioning unit, the person, as his focus (in a present environment and specific culture, with a given heredity and personal history, etc.), he might have gone further in pulling together advances in personality theory. However, while he does see the individual at times, he is usually dealing with ego as ego, an artificial abstraction representing only one aspect of the person as a unit.

The ambitious attempt, the relative success with which he channels varied contributions to his purpose, and the occasional originality in development of details make this book of interest to the sophisticated reader with time to read it. More important, Ausubel may be expected to make further significant contributions to personality theory in the future.

Roy M. Hamlin
Western Psychiatric Institute and Clinic
University of Pittsburgh

HOOKER, DAVENPORT. *The prenatal origin of behavior*. Kansas: Univer. of Kansas Press, 1952. Pp. viii+143. $2.50.

This small volume is based on the Porter Lectures delivered at the University of Kansas. Its author is one of the world's most distinguished and original students of the development of behavior. He summarizes the early growth of activity in infrahuman vertebrates and the sequence of fetal activity in human embryos and fetuses. He gives special consideration to the neural basis of the responses which he describes.

The author's own experimental work leads him to the conclusion that the behavior of vertebrates, including man, has its beginning in extroceptively initiated responses. External stimulation, it is further shown, releases a succession of responses in various vertebrate organisms which are basically similar. In general, he favors the view that behavior is individuated out of a more general matrix rather than being the integration of previously more simple reactions. All relevant experimental work is carefully reviewed in connection with this problem. Especial attention is given to the points of view of Coghill and Windle.

Hooker is notable among those who have studied the early development of behavior because of his careful description of the stimuli which he has used. In his discussion of Carmichael's work reference to the stimuli used is also given in some detail. The fact that Carmichael used the method of von Frey in calibrating hairs in units of tension in which both pressure and radius are considered is not mentioned, however. This may seem to be a small point, but an agreement upon the best form of tactile stimuli to be used

in behavior work is not without importance.

The role of the endocrine organs in fetal life is considered. An interesting diagram recently developed by Humphrey is given of the probable reflex pathway for contralateral neck and upper-trunk reflexes in the human embryo.

The author has performed a service for all students of behavior, and especially for psychologists, in bringing together in a clear and concise manner much of the scientific material which bears on the origin and development of behavior. The reviewer feels that the present volume should be in every psychologist's library.

LEONARD CARMICHAEL

Tufts College

MIKESELL, W. H., & HANSON, G. *Psychology of adjustment*. New York: Van Nostrand, 1952. Pp. ix+406. $4.50.

This is a text designed for those courses which are growing in popularity and titled variously, but often called Personal Adjustment or Mental Hygiene. The expansion of these courses can probably be attributed largely to students' demands for assistance in understanding themselves and directing their lives, and to the growth in colleges of group approaches to personal guidance. In evaluating the recent texts in this area several factors must be considered: introductory psychology texts are expanding the content devoted to personal adjustment; there is a growing literature on individual and group therapy and on student-centered teaching, which has a very direct bearing on this psychological self-help project; moreover, psychologists as scientists have turned a properly critical eye on the handling of this relatively new service; last,

it seems that we are in the midst of a transition in our goals and methods of teaching this kind of material.

It appears then that the task of writing a personal adjustment text today is a difficult one. A new text in this area should be expected to make a contribution on the subject beyond that offered by the best introductory volumes. It should lend itself to newer methods of student-centered teaching by making good didactic treatment secondary to planned stimuli for self-discovery and personal development. This would also include the inculcation of more realism, some training of the student to recognize and resolve basic life conflicts, and a gradual substitution by him of effective outlets for his self-defeating habits and attitudes.

Where can the Mikesell and Hanson text be placed in respect to these difficult goals? The table of contents suggests a systematic organization. Chapters on basic concepts such as frustration, conflict, and "wants" launch the book. There follow seven chapters in common problem areas: sex, education, work, family. The remaining chapters, with topics such as therapy, choice of mechanisms, and confidence, have a therapeutic thread running through them. Most of the concepts introduced are defined in terms of several currently used text books.

This apparently systematic treatment breaks down as the book is examined more closely. Drive, urge, need, force, and motive are used in various places without indication of the authors' evaluation of them as synonymous or different aspects of the total motivational process. The authors do not distinguish between drive and instrumental act in emphasizing urge-mindedness rather than

goal-mindedness as a hygienic set. This emphasis on drives and needs rather than on final goals is good, but it is not applied often in the book. Renunciation and direct living, for example, are advised without much emphasis on how they are motivated and achieved. Defense mechanisms are not evaluated well enough in terms of the contexts in which they may be found. Hobbies, reading, writing, and occupation are evaluated as withdrawal mechanisms without fully emphasizing the conditions under which these same outlets may be creative. The relationship between the concepts used is not always clear. One of the main characteristics of maladjustment presented is "misfitness" which seems tautologous. Maladjustment is characterized in terms of symptoms instead of personality structure.

In discussing their objective the authors claim that they will introduce theory only to bring out practical points. They purport to present facts about and remedies for the errors of human living, yet usually several definitions of a basic concept are given instead of a simple one favored by the authors. Nine classifications of "wants" are presented. In fact, there is considerable unevenness throughout the book in the writers' attempts to be academic or practical. For example, parts of the chapter on therapy appear to have been written from a descriptive and academic standpoint rather than with the view of arousing therapeutic practices in the reader.

The book contains much that the authors have undoubtedly found helpful to students taking their courses, when it was presented in an effective teacher-student context. On the whole the text is quite traditional

in its approach. It does not differ greatly in approach from many existing texts in the field.

FRED McKINNEY
University of Missouri

BERNARD, HAROLD W. *Mental hygiene for classroom teachers.* New York: McGraw-Hill, 1952. Pp. xiii+472. $4.75.

Bernard considers mental hygiene as a point of view which requires that education provide the most adequate satisfaction of and a minimum of conflict between the fundamental human needs of all individuals. He clearly demonstrates how all aspects of the school, the curriculum, teaching methods, "discipline," and grading have a bearing upon mental hygiene. Constructive and more healthy procedures are recommended to replace the many familiar and traditional ways of doing things which are inimical to good mental health. The teacher's mental health, as the principal factor influencing adjustment of pupils, is given extensive treatment.

Mental hygiene is also represented as a set of techniques. One of the major purposes of this text is to make these techniques available to every teacher, since there are many books for use of the specialist and for those "concerned with guidance as an approach" to mental hygiene. As Bernard claims, the services of the specialist are not available to many schools, particularly smaller ones. While respect for the psychologist and psychiatrist is shown, the student using this book might get the impression that such specialists are not really needed. The teacher is told not to expect to succeed with all children because adjustment is influenced by factors other than the school, but the fact that a specialist might succeed where the classroom teacher fails is not mentioned in this connection. The reference to the "formal guidance approach" indicates an alignment with many educators who fail to see that a school which is organized to meet the needs of the child has the guidance approach whether it is formalized or not. It is safe to say that no school can fully meet the needs of all children without the help of specialists, whether they be psychologists, psychiatrists, or guidance counselors. One gets the impression that the specialist may perhaps deal with the "abnormal" child, but when the teacher asks, "How can I tell if a child is abnormal?" the answer seems to be that "it all depends."

Perhaps the most confusing statements are those having to do with marking. Presumably mental health can be fostered better by "evaluating" than by "marking." This may be true if one assumes, as Bernard seems to do, that there can be only poor marking and that the faults of the curriculum may be attributed to marks. Bernard would like to see children evaluated in terms of their capacity, but never considers what will happen when the dullard who "has done as well as could be expected" finds that he has not done well enough to be the physician or engineer he aspired to be. *One important need for mental health is the fortitude to face reality.*

Considerable space is devoted to the use of art, literature, play, and drama as techniques for understanding children and providing avenues of expression and personal fulfillment. ". . . the classroom teacher can use any or all of these approaches if he has only the willingness to experi-

ment." But more is needed than a yen for experimentation, which seems to mean "try it and see what happens." Of course, what Bernard quite properly means is that the teacher should be willing to break with tradition, but little psychological foundation is given in terms of which the teacher may judge the value of what happens when innovations are tried.

Very little reference is made to research literature but there is wide quotation from reputable authorities. Further readings are suggested at the end of each chapter, many of which are hortatory. Some old references might well be replaced by newer ones. Study questions and visual aids are listed for each chapter.

The psychologist could well use this book in a supplementary fashion in courses for teachers because of its close touch with their real problems. Its use by an instructor without adequate psychological preparation and clinical experience, however, might well lead to the belief that fostering mental health is much simpler than it is.

S. S. MARZOLF
Illinois State Normal University

ANDREWS, K. R. *The case method of teaching human relations and administration.* Cambridge, Mass.: Harvard Univer. Press, 1953, Pp. xvi+271. $4.50.

ANDREWS, T. G. (Ed.) *Méthodes de la psychologie.* (2 vols.) Paris: Presses Universitaires de France, 1952. Pp. vii+882. 1,500 fr., ea. vol.

BROWER, D., & ABT, L. (Eds.) *Progress in clinical psychology.* Vol. I, Sec. 2. New York: Grune & Stratton, 1952. Pp. xxiii+564. $5.00.

CABOT, HUGH, & KAHL, JOSEPH A. *Human relations.* Vol. I. *Concepts.* Cambridge, Mass.: Harvard Univer. Press, 1953. Pp. xxxi+333. $4.75.

CABOT, HUGH, & KAHL, JOSEPH A. *Human relations.* Vol. II. *Cases.* Cambridge, Mass.: Harvard Univer. Press, 1953. Pp. viii+273. $4.25.

Contributi del laboratorio di pischologia. Vol. XL and Vol. XLI. Milan: Societa Editrice "Vita E Pensiero." Pubblicazioni Dell'Università Cattolica Del Sacro Cuore. Industrie Grafiche Amedeo Nicola & C., 1952. Pp. viii+372. 2300 Lire, net. (XLI), 1600 Lire, net. (XL).

DAVIDSON, A., & FAY, J. *Phantasy in childhood.* New York: Philosophical Library, 1953. Pp. viii+188. $4.75.

EDUCATIONAL RECORDS BUREAU. *1952 fall testing program in independent schools and supplementary studies.* New York: Author, 1953. Pp. xii+76. $.150.

EYSENCK, H. J. *The scientific study of personality.* New York: Macmillan, 1952. Pp. xiii+320. $4.50.

FORTI, E. *L'émotion, la volonté et le courage.* Paris: Presses Universitaires de France, 1952. Pp. xii+262. 900 fr.

FRANKEL, G. W. *Let's hear it!* New York: Stratford House, 1952. Pp. 63. $1.00.

GARRETT, H. E. *Statistics in psychology and education.* New York: Longmans, Green, 1953. Pp. xii+460. $5.00.

GARRETT, J. (Ed.) *Psychological aspects of physical disability.* Washington, D. C.: Federal Security Agency Office of Vocational Rehabilitation, 1952. Pp. vii+195. $.45.

GELDARD, FRANK A. *The human senses.* New York: John Wiley, 1953. Pp. x+365. $5.00.

HARROWER, M. *Appraising personality.* New York: W. W. Norton, 1952. Pp. xvii+197. $4.00.

KRAFT, VICTOR. *The Vienna circle.* New York: Philosophical Library, 1953. Pp. xii+209. $3.75.

MARTIN, W. E., & STENDLER, C. B. *Child Development.* New York: Harcourt, Brace, 1953. Pp. xxii+519. $4.75.

MCFARLAND, ROSS A. *Human factors in air transportation.* New York: McGraw-Hill, 1953. Pp. xv+830. $13.00.

MOUSTAKAS, C. E. *Children in play therapy.* New York: McGraw-Hill, 1953. Pp. ix+218. $3.50.

NEWSOM, C. V. *A television policy for education.* Washington, D. C.: American Council on Education, 1952. Pp. xx+266. $3.50.

NISBET, R. A. *The quest for community.* New York: Oxford Univer. Press; 1953. Pp. ix+303. $5.00.

ROSEN, JOHN N. *Direct analysis—selected papers.* New York: Grune & Stratton, 1953. Pp. vii+184. $3.75.

SARASON, S. B. *Psychological problems in mental deficiency.* (2nd Ed.) New York: Harper, 1953. Pp. x+402. $5.00.

SKINNER, B. F. *Science and human behavior.* New York: Macmillan, 1952. Pp. x+461. $4.00.

STONE, C. P. (Ed.) *Annual review of*

psychology. Vol. IV. Stanford: Annual Reviews, Inc., 1953. Pp. ix+485. $6.00.

THOMPSON, C. B., & SILL, A. P. *Our common neurosis*. New York: Exposition Press, 1952. Pp. xxxii +210. $3.50.

UNITED NATIONS. *Preliminary report on the world social situation*. New York: U. N. Dept. of Social Affairs, 1952. Pp. v+180. $1.75.

VESTAL, P. A. *Ethnobotany of the Ramah Navaho*. Reports of the Ramah Project. Report #4. Cambridge, Mass.: Peabody Museum of Archaeology and Ethnology, Harvard Univer., 1952. Pp. ix+94. $2.50.

WALTERS, A., & O'HARA, K. *Persons and personality*. New York: Appleton-Century-Crofts, 1953. Pp. xvii+678. $4.75.

WISDOM, JOHN. *Philosophy and psycho-analysis*. New York: Philosophical Library, 1953. Pp. vi+282. $5.75.

WOLFF, W. (Ed.), & PRECKER, J. A. *Success in psychotherapy. Personality monographs*. Vol. III. New York: Grune & Stratton, 1952. Pp. viii +196. $4.75.

Psychological Bulletin

TESTING FOR PSYCHOMOTOR ABILITIES BY MEANS OF APPARATUS TESTS

EDWIN A. FLEISHMAN

USAF Air Training Command
Human Resources Research Center[1]

One area of aptitude testing which has received relatively little development is that of psychomotor performance. Yet, it appears that a number of job specialties involve sizable components of motor activity. While the field of motor skills *learning* is a thriving basic research area, there has been only a limited attempt to examine the problem of motor skills from the point of view of *aptitude testing*. Moreover, basic research in this aptitude area has been quite limited compared with research in other aptitude areas. Even in the Air Force research program, which probably represents the most ambitious program of psychomotor test development ever attempted, efforts at a basic research level have lagged far behind test developments in the other fields.

A thorough picture of the Air Force program of research in perceptual-motor skill testing during World War II has been presented by Melton (18). The present paper will touch only briefly on some of the problems and findings discussed there. Although previous research in this area will be explored, primary emphasis

will be given to a discussion of some of the fundamental problems and issues involved in motor skills testing research and to implications for possible future test development and research. A major section is devoted to a review of previous factor analysis studies in the area of psychomotor performance.

It will be useful at the outset to distinguish between certain concepts in this area about which there has often been some confusion.

In the field of motor ability testing, the subject is generally presented with some standardized task in which he must respond by means of certain muscular activities rather than by some verbal means. The primary interest is in individual differences with respect to these *response aspects* of the subject's behavior in the task situation. The term "motor" refers primarily to the muscular activities which can be measured. However, it is true that in these tasks the subject is responding to some simple or complex stimulus situation; hence, the term "perceptual-motor" or "psychomotor" is often used. The distinction between perceptual and motor skills is a somewhat arbitrary one, each class of skills being represented in varying degrees in the performance of different tasks. There is still no general agreement concerning whether it is more desirable to con-

ceive of tasks as varying along a single dimension from perceptual to motor or as varying along two dimensions, complexity of stimulus and complexity of motor activity required. The distinction between perceptual skills and motor skills in this paper is based only upon the relative emphasis given to certain factors in the performance situation. The concern will be with tests measuring individual differences in the primarily motor aspects of the test situation. Thus, the *making* rather than the *selecting* of the response will be the important consideration, although the kind or intensity of motor activity may be influenced in varying degrees by the nature of the stimulating conditions. The use of the term "psychomotor" is a recognition of this latter fact. For example, one may be interested in measuring the subject's coordination or speed of certain motor responses with his perception of certain cues. In all cases, the primary interest of this paper is in the motor activity involved.

HISTORICAL FRAMEWORK OF PSYCHOMOTOR TEST DEVELOPMENT

The use of tests of psychomotor functions is as old as the history of individual differences measurement itself, dating back to the work of Galton and James McKeen Cattell in the 1890's. At about the same time, Munsterburg, Jastrow, Kraepelin, and others were including simple motor skills tests in their investigations of "mental" ability. Complex mental processes were believed to be best understood by analyzing them into their elementary components, usually of a sensory-motor nature. However, starting with the work of Binet and Henri, and Ebbinghaus, the trend was toward the development of more complex tests of "intelligence." Although certain motor tasks were included in the early Binet scales, the development of "intelligence" tests and motor skill tests went their separate ways. Intelligence testing, aided by group testing development in World War I, attained undreamed-of proportions. When many exaggerated initial expectations became unfilled, there was a shift in emphasis from the exclusive use of intelligence tests to the measurement of *special* aptitudes, especially with adults and older adolescents. Special combinations of tests in different aptitude areas into test batteries, were now used for classification and selection purposes. With the possible exceptions of certain dexterity tests, psychomotor tests were almost totally excluded from these batteries. Perhaps the most significant development in the trend away from complex general intelligence tests was the refinement and application of factor-analysis techniques. New insights into more basic and independent categories of ability had emerged, and tests could be constructed to sample, in relatively pure form, each separate ability area. The fruitfulness of this development was even further enhanced when the factor composition of jobs and criteria could be included in the factor analysis (13, 19).

Test development in the field of motor skills has not exactly paralleled this development of tests in the more "intellectual" areas. Early tests of motor skills were of the simplest kind. Much of the research on these tests was confined to the laboratory situation and was designed primarily to investigate such things as generality versus specificity of simple motor abilities, and also the factors underlying individual differences in certain motor skills. Thus, studies by Robert Seashore and his co-workers (27, 30, 34), Reymert (22), and Camp-

bell (5) indicate that in fine motor skills the sense employed is of moderate significance, the musculature employed is of very slight significance, and the pattern of movement involved is likely to be the most important factor. However, a basic classification of these patterns is still awaited (this is further discussed in a later section). Moreover, the investigators largely concluded that motor factors are relatively few and very narrow in scope (4, 20, 21, 25, 27, 28). The studies have generally shown simple motor skill tests to correlate very low with each other.

Under Seashore's influence, a number of simple motor skill tasks were standardized and tried out on on-the-job criteria. The Stanford Motor Skills Unit (24) was an example of these tests. Several validity studies using these tests were subsequently made. Examples are those of Harney (described by Seashore [28]) in predicting high school shopwork grades, by Walker and Adams (40) in predicting typewriter skill, and by S. H. Seashore (33) in predicting success in winding-machine apprentices in a knitting mill. The validities in all cases were insignificant. Moreover, other studies (25) showed low relationships between total training that people had in certain manual skills and many tests of fine motor skills. There have been, however, some notable exceptions. These include a study by Spaeth and Dunham (35), who reported a correlation of .61 between a test for precision in thrusting a stylus at a graded series of holes and target rifle shooting for 73 Army men. Seashore and Adams (29) also found that five simple steadiness tests distinguished sharply between a university rifle team and 50 unselected ROTC students. Humphreys, Buxton, and Taylor (17) corroborated these findings in a similar study.

They found a correlation of .77 between similar steadiness tests and rifle marksmanship.

Studies of finger and manual dexterity tests have also on occasion shown some validity for watchmakers, electrical fixture and radio assemblers, coil winders, packers and wrappers, and certain kinds of machine operators. Summaries of such validity studies involving simple dexterity tests can be found in Tiffin (39) and Super (36).

The assumption underlying simple motor skill test development was that it should be possible to develop a battery of simple motor tests which would indicate likelihood of success in a more complex motor skill. So strongly was this belief held that failures in prediction have often been attributed to faulty techniques such as lack of reliability of the measures used. In some cases this was probably justifiable, especially since in most of the studies the reliabilities were not determined, or, at least, not published. However, in other studies in which reliabilities were known (.80 to .90) the conclusions were the same.

Complex tests of motor skills were not developed to any extent until the World War II Air Force research program. That these kinds of tests had substantial validity was demonstrated. Here, paradoxically, were two trends going in opposite directions. In the paper-and-pencil test areas, test development was increasingly being aimed at simpler tests of one factor at a time. Factor analysis of printed tests was continuously utilized to obtain fewer tests, as pure factorially as possible. On the other hand, psychomotor tests were becoming more complex using increasingly complicated apparatus. These apparatus tests, although contributing considerable validity, correlated quite highly with each other. More-

over, there was no concerted effort, as was made in the paper-and-pencil area, to use factor analysis data in developing and organizing apparatus tests along new and relatively independent dimensions.

SOME CHARACTERISTICS OF APPARATUS TESTING

Recent aptitude testing, in the area of perceptual-motor ability, to a great extent has consisted of individual apparatus tests. The apparatus may vary in complexity from simple pegboards to complicated mechanical or electronic contrivances. The assumption underlying the development and use of apparatus tests is that certain kinds of abilities can best be measured by these performance tests as contrasted with paper-and-pencil tests. Individual apparatus tests often appear necessary whenever the primary interest is in the motor aspect of the subject's responses. Such functions as perceptual-motor coordination, smoothness of control movement, speed of discriminative reactions, appropriateness of control movement, responsiveness to kinesthetic cues, and motor control under stress conditions are examples of skills for which apparatus tests seem more suited than printed tests. Whenever the interest is in the *making* rather than the *selecting* of the response, some instrument is usually needed to provide the cues to be responded to, the means of response, and the means for recording the speed, precision, or other relevant features of the response.

Very detailed descriptions of apparatus tests developed in the Air Force Classification Program may be found in Melton (18). Some of the better-known tests described include the Complex Coordination, Two-Hand Coordination, Rotary Pursuit, Rudder Control, Discrimination Reaction Time, Aiming Stress, and Finger Dexterity Tests.

In most cases, each test consists of a task unit (which the subject operates) and a control unit which contains the timing apparatus, counters which record the scores, and switches which the examiner uses to control the testing period. There may be from two to ten task units (depending on size) connected to one control unit.

Apparatus Test Problems

Melton (18) has summarized the problems that accompany the use of apparatus tests and the steps that were taken to solve them in the Air Force program. Some of these difficulties are sketched only briefly here.

First, such tests are expensive to build and expensive to maintain. Moreover, since a test can be given to only a few subjects at a time, apparatus tests are administratively more expensive than tests which can be given to groups of a hundred subjects by only three or four administrators. The task of maintaining uniformity of testing conditions for each man tested is also increased with apparatus tests. This includes lack of uniformity owing to examiner differences and to apparatus differences. Examiner differences can be reduced by increased training. There is also some evidence (2) that the examiner is not a very potent source of variability, at least in situations where the possibility of timing and recording errors is minimized. Maintaining uniformity of the apparatus from person to person is a more serious problem. This includes maintaining comparability of scores within the same piece of apparatus as well as differences between several copies of the same apparatus. In the Air Force program, it was found necessary to

build apparatus to high standards of precision and to subject them to a rigorous program of preventive maintenance and of calibration. In addition, a system of statistical control to check on apparatus differences was also found necessary.

Several other apparatus testing problems have received consideration. Studies investigating whether the number of subjects tested at one time influenced individual test scores have generally shown little effect of this "social-interaction" variable (1, 18).

In another study, it was found that different orders of presentation of different apparatus tests in a battery affected scores significantly in the case of only one test (Aiming Stress). However, it would be hazardous to extrapolate to other test batteries from this generally negative finding. At any rate, some constant order of presentation of each test relative to other tests in a battery would seem to be necessary.

Apparatus tests also present special problems in the determination of intratest and test-retest reliability coefficients, since performance of subjects in the tests almost invariably shows improvement with practice. The usual procedure is the intratest correlation where the total test period is divided into trials and odd versus even trials are correlated. The chief value of these coefficients is in their use relative to the interpretation of intercorrelations with other tests in the battery given at the same time.

Thorndike (37) has pointed out an additional difficulty encountered in using apparatus tests. This is the difficulty of assembling test records fast enough that validation data can become available upon an adequate sample within a reasonable time. Because only a few pilot copies of such tests are usually available for such experimental use, testing may have to continue over long periods of time in order to accumulate sufficient samples for whom criterion data will become available in different job specialties. Adding the time necessary for the criterion data to mature, it can be seen that the validation program for apparatus tests can be a slow-moving affair. The present Air Force program has partially solved this problem by the use of mobile trailer units containing copies of the tests. These units can be moved around the country to different Air Force bases whenever desirable criterion groups are available.

Although it is important to take into account these limitations of apparatus tests, and to hold the controllable sources of error to a minimum, the value of the tests must ultimately be assessed in terms of the unique contribution they can make to the over-all prediction of the criteria. That "something" unique (not tested by other tests) has been added to the Air Force batteries by apparatus tests has been repeatedly demonstrated (13, 18). For example, the addition of six psychomotor tests to fourteen printed tests raised the multiple correlation of the battery with the pilot criterion from approximately .50 to .70.

Factor analyses of the complete battery consistently revealed three psychomotor factors measured by the apparatus tests included. A factor identified as "Psychomotor Coordination" was found to be quite general. It was found in tests requiring small-muscle adjustments (finger dexterity tests) as well as tests requiring large-muscle adjustments (movements of arms, torso, and legs). A second factor was tentatively named "Psychomotor Precision" because it was involved in tests requiring accurate

manipulations under speed conditions (discrimination reaction time, finger dexterity). A third factor was tentatively called "Psychomotor Speed" because it was primarily involved in tests requiring sheer speed of marking an answer sheet. An additional factor hypothesized was described as a "Kinesthetic-Motor" factor measured almost uniquely by the Rudder Control Test. The analyses also indicated substantial loadings of these factors in the different criteria. The factor analyses also revealed, however, that a few paper-and-pencil tests in the battery sampled at least some factors measured by apparatus tests. Moreover, the indication from the test communalities was that the variance of the psychomotor tests in general is not as well accounted for by the present factors as that of the printed tests.

Implications of these findings and more complete descriptions of factor analysis studies of motor skills tests are presented in later sections.

Factor Analyses of Psycho-motor Tests

It seems apparent that some extensive dimensional analysis of motor abilities is necessary. It is also proposed that factor analysis techniques can aid greatly in this respect. Although factor analysis methods have been employed in this area on a limited scale, they have been performed on either a limited number of simple motor tasks or on a limited number of complex tasks. A factor analysis of a much wider variety of motor tasks which could reveal a more basic classification of factors primarily important in the performance of the tasks would seem desirable. This would not only help bring order into the field of psychomotor testing but would provide a framework for research in general in the area of motor

skills. Such a methodological approach could lead to a clearer delineation of unique psychomotor factors and a better understanding of the variables that contribute to the validity of both simple and complex tasks.

In this section, a summary will be given of some *previous* factor studies that have been made. This may provide additional background for needs in this area and may also provide suggestions for the kinds of tasks to be included in future factor analysis batteries.

Difficulties in Comparing Different Factor Studies

There are definite problems and limitations which one encounters in comparing different factor analysis studies. A major difficulty is the different interpretation given to factors by different researchers. This is sometimes only a question of semantics, but in at least a few cases two investigators might give entirely different meaning to factors. Also, the interpretations given factors are often based on much too limited evidence.

There is also the difficulty of comparing factors derived by different factor methods. This paper will be confined to studies using variations of Thurstone's centroid procedure. Investigators may still differ in (*a*) their criteria of when the solution is reached, (*b*) whether they favor oblique or orthogonal solution, and (*c*) whether they rotate "blindly" or with use of certain "hypotheses." However, results generally show quite close agreement between these different emphases in rotation.

Another major problem is the lack of identity of tests in the different batteries factor analyzed. Thus, factors based on different sets of tests may be named the same but that they

are operationally the same may not have been demonstrated.

Then there is the matter of sampling errors. No satisfactory measure of the standard error of a factor loading has been determined. Thus small loadings may not indicate that the test is measuring a given factor. However, if these small loadings repeatedly show up on replications of the factor analyses, one might have considerable confidence in their significance.

The summary of previously isolated psychomotor factors to be presented here is not intended to be as detailed as the excellent surveys published by Cattell (6) in the field of personality or by French (9) in the field of aptitude and achievement testing. However, it is intended to be comprehensive in at least the coverage of previously published studies.

Factors Derived

The studies considered here ranged from factor analyses of very simple motor tests to analyses of complex apparatus test batteries. In general, studies of simple motor skills suggest that if there are underlying factors of motor ability they are relatively narrow and do not extend over a wide range of tests. In more complex tests broader factors appear. Some of the *nonmotor* factors which seem most consistently identified in psychomotor test batteries will also be discussed. Of course, these latter factors can only be identified in those few cases where nonmotor tests were included in the total test battery.

Reaction Time. This refers to the speed with which an individual can make a predetermined response, usually of a fairly simple type, to a presented stimulus. This has appeared repeatedly as a separate factor wherever measurement of rela-

tively simple reaction time tests has been included in the analysis. Thurstone (38) in his analysis of many tasks loaded with perceptual elements found a factor common only to reaction time to light and reaction time to auditory stimulation. Similarly, Seashore, Buxton, and McCollom (30) identified such a factor with loadings on visual jump (subject has to move his hand six inches to the peg) reaction time, auditory simple reaction time, visual simple reaction time, simple horizontal tapping, and auditory jump reaction time. The appearance of this factor emphasizes again that the pattern of movement is of greater significance than the particular sense modality or musculature involved. These studies have involved simple reaction time tasks. There is some evidence that this factor may extend to more complex reaction time tasks. Thus, Seashore, Starman, Kendall, and Helmick (32) found correlations of from .63 to .98 among reaction time tasks of varying complexity.

Tapping Ability. This is the speed with which the subject can oscillate either his fingers or his arm. This seems relatively independent of any eye-hand coordination ability. For example, Greene (12) found this factor in his analysis of a variety of tasks involving aiming, tapping, and dotting with telegraph keys, fingers, pencils, etc. He found the more eye-hand coordination involved in the task, the smaller became the loadings on the Tapping factor. Thus, tapping with a pencil, making no effort to tap in a particular spot, yielded a high loading on Tapping. Tapping in large circles yielded moderate loadings on Tapping and similar loadings on another factor he named "Aiming." Finally, tapping in small circles where positioning the pencil was difficult yielded practically no

loading on Tapping but high loadings on Aiming. The Aiming factor is discussed later, but it appears that the two factors are usefully considered separate. Tests which Greene found with loadings on this Tapping factor included: dotting in large circles and tapping with either the left or right hand. Guilford in his factor analysis of dexterity tests (reported by Melton [18]) identified Tapping as a separate factor. The only two tests with appreciable loadings on this factor were tapping using finger action and tapping using wrist action. Similarly, Wittenborn (41) found this factor in his analysis of mechanical ability tests (although he called it "Repetitive Movement"). Seashore, Buxton, and McCollom (30) also isolated a Tapping factor in their test battery but they further broke it down into two subfactors. One seemed to involve finger-hand speed in restricted oscillatory movement in one place only. Tasks using tapping keys, a two-finger oscillometer, and short movements had loadings on this factor. The second factor involved forearm and hand speed in oscillatory movements of moderate extent in two planes. This factor was best measured by stylus tapping on two or three plates successively. These two subfactors isolated by Seashore, Buxton, and McCollom may be related to the finger versus manual dexterity distinction discussed later.

Psychomotor Coordination. This factor has appeared in all the Air Force analyses which included apparatus psychomotor tests. It represents either integration of muscular movements or coordination between the eye and muscular movements. It is measurable in the rotary pursuit, complex coordination, two-hand coordination, finger dexterity, aiming stress, and rudder control tasks. It is thus quite general to skeletal musculature and, though common to finer as well as grosser movements, seems best measured by movements of moderate scope. Analyses revealing this factor have been discussed in the Air Force research reports (13, 18), by Dudek (8), Michael (19), and by Guilford and Zimmerman (14).

Zachert and Shibe (42), in their factor analysis of the United States Employment Service Battery, which contained no complex apparatus tests, identified a factor as Psychomotor Coordination. The assembling, disassembling, and placing blocks tests had the highest loading on this factor. However, it appears that the factor is the same as what has usually been called a Manual Dexterity factor. With present evidence, it seems safer to consider Manual Dexterity separate from Psychomotor Coordination. Tests of the factor Psychomotor Coordination involve more coordination between muscle groups, are not entirely restricted to arm movements, and do not seem as concerned with speed. It is possible that the Psychomotor Coordination factor may break down into several simple factors if other kinds of less complex tasks were also included in these factor analysis batteries.

Manual Dexterity. This factor involves arm-hand coordination and speed. Tests with loadings on this factor require skillful, controlled arm or hand movements at a rapid rate. The factor appears in analyses of batteries such as the Minnesota and United States Employment Service Tests. Wittenborn (41) found this factor in a battery containing many variations of pegboards. This study also indicated that this factor does not depend on the visual sense. Thus, a "pegs" task, where there was a screen in front of the pegs, and a "tactual" task of sorting by touch

revealed loadings on the factor. The United States Employment Service analyses[2,3,4] also showed such a factor common to tasks requiring blocks to be turned over, blocks to be placed in holes, pegs to be moved, turned, inserted, and washers and rivets unassembled. Harrell (15), in his analyses of various mechanical and manual ability tests, also isolated such a factor which he calls "Agility." Again, various pinboards, pegboards, simple assembly and disassembly tasks emerged clustered on one factor.

Finger Dexterity. Difficulty arises in distinguishing between what some writers call Manual and others call Finger Dexterity. It appears, however, that Manual Dexterity is the more general factor and may be common to tasks in which an additional factor Finger Dexterity is present. Finger Dexterity involves the rapid manipulation of objects with the fingers. It is distinguished from Manual Dexterity in that it does not include arm motion (although some tasks may require both factors). Thus, placing and turning blocks generally have loadings on Manual Dexterity, but not Finger Dexterity, and pegboards often include both factors. Guilford (reported by Melton [18]), in his analysis of dexterity tasks, found five factors to account for the intercorrelations of the 18 tests in the battery. One of these factors he named "Dexterity," and this factor included various pegboards,

marking, dotting, simple manipulations, and a finger dexterity test. His analysis did not isolate separate Finger and Manual Dexterity factors. However, many other investigators have isolated a Finger Dexterity factor and it seems useful to handle them separately although they appear related. It is also to be distinguished from Aiming in that accurate eye-hand positioning seems not to be required.

This Finger Dexterity factor has been named in several United States Employment Service test battery analyses.[5] Tests with loadings on it in one analysis included finger dexterity—assembling, and finger dexterity—disassembling (these two tests together with various pegboards also had loadings on Manual Dexterity). In other U.S.E.S. analyses, these two tests, tweezer dexterity, and two small pegboards had loadings on this separate factor in addition to loadings on a Manual Dexterity factor.

Psychomotor Precision. In several Air Force analyses (13) a factor named Psychomotor Precision seems similar to Finger Dexterity, although more eye-hand coordination seems involved. The factor was most heavily weighted in psychomotor tests requiring manipulations under speed conditions. It is distinguished from Psychomotor Coordination in that grosser arm motion is not included. Tests most constantly loaded on this factor were a "twisting pegs in a pegboard" task, the Discrimination Reaction Time Test, and in a later analysis, the Rotary Pursuit Test. This cluster has also appeared in the postwar Air Force analyses. The factor requires further investigations with new tests to clear up its status, especially in relation to Finger and Manual Dexterity. Thus, Zachert and Shibe (42), in a recent

[2] UNITED STATES EMPLOYMENT SERVICE. Factor analysis studies; report for Group 3. The preliminary factor analysis study. Unpublished manuscript, Washington, D. C.

[3] UNITED STATES EMPLOYMENT SERVICE. Factor analysis studies; report for Group 3. Unpublished manuscript, Washington, D. C.

[4] UNITED STATES EMPLOYMENT SERVICE. Report of factor analysis for Group VII. Unpublished manuscript, Washington, D. C., 1944.

[5] See footnotes 2, 3, and 4.

analysis, named one of their factors "Psychomotor Precision" with an alternate name of "Finger Dexterity." The battery, however, contained no complex apparatus tests, and the tests loaded on this factor (turning blocks, mark-making, placing blocks, disassembling) indicate that Finger or Manual Dexterity would be a more appropriate name on the basis of present evidence.

Although this factor has been named Psychomotor Precision, it seems to involve speed to a great extent. More clearly, this "precision" factor does not refer to the kinds of precision which minimize speed (such as accurate tracing, thrusting a stylus accurately in holes or holding it stationary in holes). Nor does it refer to the various types of postural steadiness involving grosser musculature. Moreover, it does not seem related to the Aiming factor which seems to involve both a series of precise movements *and* speed. At best, very little on the positive side is known about this "Psychomotor Precision" factor as tentatively identified on the Air Force tests.

Steadiness. This factor has been discussed by Seashore and his co-workers (28). The indication is that coordinations emphasizing accuracy (precision or steadiness) while minimizing speed and strength tend to cluster together. Studies bearing on this factor include those of Buxton (4), Spaeth and Dunham (35), Humphreys, Buxton, and Taylor (17), and Seashore, Dudek, and Holtzman (31). Tests usually included in this factor are tremor, stylus thrusting at holes, stylus held stationary in holes, and steadiness in tracing in a narrow V slot. The study by Seashore, Dudek, and Holtzman (31) on Arm-Hand Precision Tests tested the possibility of finding more than one factor in the range of tests

that emphasize steadiness in slower movements or in posture. They isolated three subfactors. These subfactors included involuntary movement of arm and hand, steadiness of movement in a restricted plane (e.g., thrusting a stylus, tracing a straight path towards oneself with a stylus), and steadiness in two- or three-dimensional space (e.g., moving a ring around a rod precisely). However, until a wider range of tests is investigated relative to these subclusters it seems safer to consider a "steadiness cluster" involving precise, slower arm-hand movements and one other grosser kind of precision cluster measured by various postural adjustments such as sitting and standing sway. The moderate relationship between these two clusters has also been shown by H. G. Seashore and Koch (23) and Seashore, Buxton, and McCollom (30). This latter factor may be like the factor of kinesthesis tentatively identified in the Air Force battery.

Motor Kinesthesis. The existence of this factor still rests upon an insecure basis although it has appeared in several Air Force analyses (13, 18, 19). The factor was almost entirely confined to the Rudder Control Test. This test consists of a simulated cockpit in which the subject has to make compensatory motor responses to keep the unit balanced in a given position. He does this by pushing rudder control pedals but the displacement of his entire body from the position of equilibrium brings into play kinesthetic and tactual factors. The test also has loadings on the Psychomotor Coordination factor previously discussed, but the two factors may be separate. Later Air Force analyses included pilot interest and pilot experience scores loaded on this factor. How this factor relates to the factor common to

standing and sitting sway identified by Seashore, Buxton, and McCollom (30) in their studies remains to be demonstrated. Including tasks of the latter type in batteries with the Rudder Control Test might clear up some of these points.

Aiming (*Paper-and-Pencil Psychomotor Speed with Precision*). This factor represents the ability to carry out quickly *and* precisely a series of movements requiring eye-hand coordination. Unfortunately, all tests yielding this factor have been of the paper-and-pencil type and there is yet no evidence that it extends to other kinds of rapid and precise manipulation. Its relationship to Psychomotor Precision, for example, is unclear. However, the United States Employment Service analyses[6] indicate it is separate from the Dexterity factors, the study by Greene (12) confirms its distinction from Tapping, and the analysis by Guilford (reported by Melton [18]) reveals it is separate from both Tapping and Dexterity. In Greene's factor analysis, the rapid placing of dots in a series of small circles with either hand yielded high loadings on this factor, but none on Tapping. Similarly, tapping in large circles or tapping in no particular spot yielded no loadings on Aiming. In the U.S.E.S. analyses, this separate factor appeared containing tasks requiring speed in dotting in small circles (again the smaller the circle, the higher the loading), rapidly making three lines in each of a series of squares, and rapid drawing of a line through large *H*'s without touching the *H*'s sides. Chapman (17), in his factor analysis of Mechanical Ability Tests, found separate Space, Dexterity, and Aiming factors. Although he called the latter "Motor Controlled Manual Movement," this factor contained

the same speeded, precise tracing and dotting tasks. Goodman (11), with the same battery and a different factor extraction and rotational system, found this same factor. Guilford (reported by Melton [18]), in his dexterity battery analysis, found speed of dotting in circles and speed of marking an IBM answer sheet pattern had highest loadings on a single factor and very low to zero loadings on the other four factors isolated. He named this factor simply "Paper-and-Pencil Motor Skill."

In several Air Force battery analyses (13, 18, 45) the name of "Psychomotor Speed" has been given to a factor including some of these same types of operations. Highest loadings were on log book accuracy (speed in marking in the indicated *A*, *B*, *C*, *D*, or *E* slots on an answer sheet) and marking accuracy (speed in marking an answer sheet under certain circled letters). The Psychomotor Speed label seems too broad a name since there are a considerable number of unrelated psychomotor factors that emphasize speed. Zachert and Shibe (42) also found this factor loaded with similar tasks in the U.S.E.S. battery. They defined the factor as "the ability to carry out a series of movements quickly and precisely," which is very similar to our definition of the Aiming factor.

Ambidexterity. This factor has been identified in only one analysis (Greene [12]), but may bear further investigation. The tests having principal loadings on this factor were aiming and tapping tests performed with the left hand. Since 90 per cent of the subjects in Greene's study were right-handed, this factor was thought to be the ability to use the nonpreferred hand. The tests with loadings were dotting in small circles, dotting in large circles, and speed of tapping —all done with the left hand. Addi-

tional studies such as ones requiring left-handed people to use the right hand and studies manipulating handedness in other tasks might help confirm or reject this factor.

Nonmotor Factors in Primarily Motor Tests

Factor studies (as well as experimental ones) indicate that success in complex motor activities may depend upon nonmotor as well as motor factors. Continuing within the framework of previously isolated factors, a few of these nonmotor factors will be discussed briefly.

Some form of *Spatial Relations* seems involved in many psychomotor skill tests. This factor appears to represent the ability to relate different responses to different stimuli, where either stimuli or responses are arranged in spatial order. Thus far, the evidence limits the identification of this factor to tasks involving visual perception only. In at least six separate analyses (13, 18, 19) this factor was found to be included in some apparatus tests. The Complex Coordination, Discrimination Reaction Time, and Two-Hand Coordination Tests were most consistently loaded on this factor, along with the paper-and-pencil nonmotor tests of Dial and Table Reading, Coordinate Reading, and Instrument Comprehension. Seashore, Buxton, and McCollom (30) found a factor they named "Manipulation of Spatial Relations" which included rotary pursuit, assembly, and form board tasks.

Mechanical Experience is another factor which consistently showed up in the Air Force analyses (8, 13, 18, 19) as including a motor task. However, the Two-Hand Coordination Test was the only motor test with significant loadings on this factor. Other tests in the factor were Mechanical Information, Mechanical

Principles, Biographical Information, Pilot Technical Vocabulary, and Reading Comprehension.

A factor named *Pilot Interest* was identified in later Air Force studies (8). This factor was separate from Mechanical Experience and included the Rudder Control Test, general information, and previous flying experience.

A *Perceptual Speed* factor has been found in complex motor tests in some AF analyses but not in others. This factor includes the ability to make rapid recognitions and comparisons of visual forms. The analyses of Dudek (8) and Michael (19) and those reported in the Air Force volume (13) have indicated Discrimination Reaction Time and/or Complex Coordination loaded on this factor. In both these tasks different light patterns are presented successively and must be discriminated rapidly by the subject before he can make the appropriate motor responses. Guilford (reported by Melton [18]) found Discrimination Reaction Time, rapid number cancellation, and an arm-hand coordination task of interchanging pegs grouped with two spatial relations tests in a factor he called "Perceptual."

Factors in the *Intellectual-Verbal* area have not been shown to include motor tasks to any extent. They did not seem to enter very much into performance on any of the complex apparatus tests. Although Guilford (reported by Melton [18]) found a dexterity test of moving round pegs forward, number cancellation, and simple conflicting manipulations to group with "following directions" and mechanical comprehension tests, his naming of this factor as "Intellectual-Verbal" should be regarded as very tentative, however.

Perhaps just as notable as the presence of some of these factors is

the *lack* of very many nonmotor factors that have been identified in motor performance. However, there was not much opportunity, with the exception of the Air Force analyses, to identify the nonmotor factors in motor task performance since few batteries included both kinds of tests. Future analyses can be aimed also at this problem. From the point of view of developing "unique" psychomotor tests, the effort should be made to minimize in such tests the nonmotor variance measurable by paper-and-pencil tests.

Factor Summary

The psychomotor factors discussed above have generally appeared in two or more factor studies. Although the status of some of these factors is still in doubt, there is considerable agreement about other factors.

The nature of seven of the ten factors discussed rests on more secure ground than the remaining three.

From the limited number of such studies that have been made thus far, the following motor skill dimensions seem to emerge:

1. The speed with which an individual is able to respond, by means of a prescribed movement, to a stimulus when it appears (Reaction Time).

2. The speed with which an individual is able to oscillate either his fingers or his arm, independent of any eye-hand coordination ability (Tapping).

3. The ability to make skillful, controlled arm or hand manipulations at a rapid rate (Manual Dexterity).

4. The ability to make skillful, controlled manipulations with the fingers at a rapid rate (Finger Dexterity).

5. The precision and steadiness with which one is able to make accurate arm-hand positioning movements which minimize strength and speed (Steadiness).

6. The ability to carry out quickly *and* precisely a series of accurately directed movements requiring eye-hand coordination (Aiming).

7. The ability to make somewhat precise postural or bodily adjustments to kinesthetic cues when the body or body members are displaced from positions of equilibrium (Motor Kinesthesis or Gross Precision).

Somewhat less well-defined dimensions seem to be:

8. The ability to integrate gross or fine movements of moderate scope. This factor was identified whenever more complex apparatus tests were used and seems, at present, to be a general muscular agility factor. (It was named "Psychomotor Coordination.") However, it is quite possible that this factor may break down into several more simple factors.

9. The ability of right-handed subjects to use the nonpreferred hand under speed conditions. This has been identified in only one study and was called "Ambidexterity."

10. A factor called "Psychomotor Precision" is an even less defined factor. It may be the same as Finger or Manual Dexterity, but thus far there has been no way of evaluating this. Tests involving this factor seem actually to require more speed than precision.

Nonmotor factors on which certain psychomotor tests have carried loadings include Spatial Relations, Perceptual Speed, Mechanical Experience, and Pilot Interest.

It may be stressed here that only *factor analysis* studies on dimensions of motor skills have been summarized in this section. There are other ways of categorizing and describing such skills. However, it is felt that factor analysis provides the best available

technique for classifying skills into the smallest number and most independent set of categories that might account for performance on a wide variety of tasks.

POSSIBLE RESEARCH APPROACHES

Factor analyses of psychomotor abilities. From the foregoing review it can be seen that much additional research is needed to clarify a dimensional analysis of motor abilities. It is the view of the present author that the major emphasis in future research in this aptitude area should be given to systematic investigations of basic psychomotor ability dimensions. In the preceding section some starting points for more definitive analyses are suggested by the previous factors identified, tentative factors needing confirmation, and the limited range and complexity of tests utilized. It would seem that the construction and factor analysis of comprehensive, specially designed psychomotor test batteries are basic needs in this area. Such research would help to get us closer to the isolation of the relevant variables in motor performance with respect to aptitude test development.

On-the-job surveys. Surveys of the specific motor skills which appear to be involved in various jobs might provide assistance in designing tasks to be included in factor analysis batteries. Various job analysis approaches might be tried, although the analyst would have only traditional motor skill categories available. As a side study, it might be possible to "job analyze" a well-controlled complex motor task along different kinds of dimensions. Simpler tests could then be constructed to measure the components thus derived. Then it might be possible to see which kind of dimensional analysis could best predict over-all performance on the complex task. Such a study might suggest some possible approaches to on-the-job surveys which would depart from traditional techniques.

However, any kind of careful job study might aid materially in the construction of a meaningful factor analysis battery of motor tasks.

Transfer experiments. Laboratory transfer experiments might be a potent source of data on functionally similar or different aspects of motor tasks. A series of such experiments which systematically varied both the stimulating conditions and the responses involved might yield evidence on common elements in a given variety of tasks. Research on transfer has generally been undertaken within the framework of studying factors influencing *learning*. However, it would seem possible to view such experiments from the standpoint of aptitude testing research, especially where transfer from one task situation to another is studied.

Other experimental investigations. Other kinds of experimental investigations involving motor tasks would give leads with respect to the significance of variables in the motor area. Manipulations of procedural, task, and motivational variables with special attention to patterns of interrelationships might throw additional light on basic factors involved. Investigations of the effects of continued practice on the performance of complex motor tasks might, for example, suggest which tasks bring into play different abilities at different stages of task performance. Or again, changes in the patterns of relationship between different tasks when one or both tasks are practiced might yield additional data on specific or common factors. Such data, for example, might help one decide at which stage of learning scores on certain tasks should be correlated with other task scores for a factor

analysis battery. On the other hand, one might want to include tasks whose correlational pattern is most resistive to change due to practice.

Research on qualitative characteristics of motor responses. This approach overlaps and perhaps should precede the on-the-job survey. It includes observation of laboratory tasks as well as actual job situations and also a literature search from a different point of view from that reported here. Here, one would hypothesize qualitatively different movement characteristics over and above those which have been involved in factor analyses or in traditional classifications. Thus distinctions have been made often between ballistic movements and tense movements, gross versus fine movements, accelerating versus decelerating aspects of movement, manipulative versus nonmanipulative movements, compensatory versus noncompensatory, etc. An example of a suggested classification of motor reactions into fairly distinct types has been provided by Brown and Jenkins (3). It might be profitable to include in any empirical dimensional analysis different tasks consisting primarily of such assumed qualitatively different responses to see how correlated they really are and just what factorial composition might emerge.

Small-scale factor analyses. The above sources, which would provide dimensional analysis data and suggestions for a more extensive factor analysis battery, overlap considerably. The collection and integration of all available information and the design of suitable tasks for the large scale analyses would take considerable time. Consequently, small-scale factor analyses on various subareas could be carried out. These analyses would also provide preliminary data relevant to the design of the more elaborate studies.

SOME METHODOLOGICAL PROBLEMS

There are some additional issues in the area of psychomotor testing which overlap somewhat with those previously discussed. Some of these will be presented briefly since they also have important bearing on the directions which psychomotor test development might take.

Job-Sample Tests Versus Tests of Basic Abilities.

Aptitude test development for the selection of personnel for a particular job often takes the form of tests that closely resemble the job both in content and complexity. The assumption here is that the more nearly the test resembles the job or some phase of it, the more accurately test performance will predict job performance. Because the job is usually complex, requiring a number of different operations often at the same time, the tests themselves become complex. This is especially true in recent psychomotor test development and was illustrated in the Air Force wartime apparatus testing program. Thus in developing tests for pilot selection a large number of complex pursuit and coordination tests were developed. Subjects were required to respond to a variety of signals and cues. Even more to resemble the job, the tests often used airplane-type stick and rudder controls and instrument panels. An additional advantage of such psychomotor tests is their high face validity for both the sponsor of the research program and the examinee. This, however, in the case of some subjects may be a disadvantage. The subject may be negatively motivated in tests which he assumes, rightly or wrongly, will qualify him over another preferred job. Such tests did, however, prove to have considerable actual

validity for predicting the particular criterion of success used. It certainly seemed that the validity of these complex tests was greater than could have been achieved by any combination of simpler tests available at the time.

However, as might be expected, each test of this type correlated quite highly with all other complex apparatus tests developed for the same job. Furthermore, as Thorndike (37) has indicated, such tests overlap not only in their valid variance, but also in their invalid variance for each job. Moreover, errors in interpreting the job also may be perpetuated in each test. The high intercorrelations indicate that relatively little gain will result from adding additional complex apparatus tests to the battery. However, as Thorndike also points out, whether the final multiple correlation will be higher for a program based on complex job analogy tests with relatively high validities and high intercorrelations or on relatively pure tests of simple functions with lower validities and lower intercorrelations remains an open question.

The issue becomes increasingly important, however, when the problem of *multiple* selection or *classification* rather than simple selection is considered. Although it would be theoretically possible to devise a separate battery for predicting success in each of several jobs to which applicants might be assigned, such an approach would become hopelessly inefficient, unwieldy, and time-consuming with a larger number of job categories. Thus, it becomes necessary to use each test for predicting success in several jobs. In the classification situation, then, it becomes less defensible to design the test in terms of a particular job. The problem is further pointed up by the fact that even the individual job itself may

change. For example, the Air Force pilot of 1960 may have to bring into play different combinations of abilities from those of the pilot of 1950.

The alternative to such job-sample tests in the psychomotor area is tests designed more in terms of general psychomotor ability categories. These tests may still be apparatus tests and may be complex mechanically and operationally, but not complex factorially; that is, each test will sample as nearly as possible only one motor "ability" category. This approach to psychomotor test development, then, starts off with the search for basic motor ability categories suggested above.

Are Complex Psychomotor Tasks Qualitatively Different from Tasks of Simpler Motor Functions?

There is a fundamental assumption underlying much of the discussion in this paper—the assumption that it is possible to break down the variance in performance of complex psychomotor tasks into simpler more fundamental psychomotor functions. As was indicated earlier, many previous attempts at predicting performance in complex motor tasks from skill in simpler tasks have failed. However, these studies did not utilize tasks representing empirically derived factors in the complex task. In the few cases where this was done (e.g., in the steadiness and finger and manual dexterity areas), predictions were more successful.

By "simple" tasks, the author does not necessarily mean such tasks as finger oscillation, peg-turning operations, or reaction-time functions, but rather, tasks which are not *factorially complex*. Thus, a task requiring a certain kind of coordination measured by complicated apparatus may turn out to be "simple" in the sense of sampling an ability area that may

vary relatively independently from another motor ability area. The apparatus complexity required for tests measuring many of the yet undefined motor skill factors may lie somewhere between the very elementary operations requiring a minimum of apparatus and the complex job-sample type tests now in use.

However, whether the introduction of complexity of function introduces valid variance not covered by *any number* of more analytical tasks of simpler motor ability functions is a point still in question. From the success of such relatively "pure" tests in other aptitude areas, it must be assumed that it is possible to predict performance efficiently in complex tasks from combinations of tests involving more basic categories of ability. Whether the abilities involved in complex psychomotor tasks involving a number of operations are qualitatively different from the abilities involved in the more analytical tasks remains a subject for future research.

Apparatus Tests Versus Paper-and-Pencil Tests of Motor Abilities

A few studies, previously mentioned, have shown that some paper-and-pencil tests have reproduced some of the variance involved in the performance of certain apparatus tests. Thus, various tests of spatial relations and perceptual speed, and the mechanical experience key of a biographical information blank were moderately correlated with some complex apparatus tests in the Air Force batteries. Tests of these kinds have been regarded as measuring nonmotor factors involved in primarily motor tasks.

On the other hand, some paper-and-pencil tests have been designed primarily as motor tests (i.e., tracing, dotting, aiming). That such tests can possibly reproduce the variance

associated with more complex apparatus tests remains to be shown. In several studies reported earlier, batteries containing both paper-and-pencil and apparatus motor tests often yielded a factor unique to the paper-and-pencil tests. Whether this is due to the restricted range of tests in the batteries or to some actual differences in motor abilities tapped by paper-and-pencil and apparatus tests needs to be demonstrated. More paper-and-pencil tests specifically designed to sample certain psychomotor functions need to be constructed and correlated with apparatus psychomotor tests. Such research might prove even more significant if the factorial content of the apparatus tests were known.

The ultimate utility of paper-and-pencil motor tests lies in their validity. There is little doubt that if ever any number of such tests can be shown to predict a certain apparatus test score efficiently they should be substituted for it.

Motion-Picture Tests

Nothing has been said thus far concerning the role of motion-picture tests in motor skill aptitude testing. This is because motion-picture tests have been used in the past to test functions primarily perceptual rather than primarily motor in nature.

The use of motion-picture tests in the wartime Air Force program has been summarized in Gibson's volume (10). General areas in which these tests were constructed included (*a*) tests of ability to judge motion and locomotion, (*b*) tests of ability to judge distance, (*c*) tests for orientation in space, (*d*) tests of ability to perceive slight movement, (*e*) tests requiring multiple perception, (*f*) tests involving sequential perception, (*g*) tests of perceptual speed, and (*h*) tests of comprehension. In all cases,

the testing emphasis was to measure accuracy of perceptual discriminations or judgments of a psychophysical nature which could not be adequately measured by printed tests. Thus, the developers of these tests were not at the time interested in measuring the *motor aspects* of responses to the stimulating conditions presented. The examinees, in fact, recorded their answers on standard IBM answer sheets.

Considerable thought might be given to the use of motion-picture projections, or series of slide projections, as an integral part of apparatus tests used in the *motor* ability area. The problem of synchronizing and recording differential response pattern with photographic stimulus presentation is a crucial and difficult one, however, and would have to be solved before such tests could be useful. One would have to investigate carefully if photoelectric synchronization could be made to meet the standards of precision demanded for proper calibration and standardization of such tests.

With the possible exception of the Pedestal Sight Manipulation Test (18), the use of photographic means of presentation has not been very much explored in the motor skill testing area. Such presentation might allow, for example, a more economical means of testing for motor skills with larger numbers of subjects at a time. Moreover, such stimulus presentation might be more flexible than one involving more complex apparatus and lighting systems. Thus, one could easily vary the speed of presentation, exposure time, and the intervals between presentations. Or, again, one could keep the mode of response constant and change the stimulating conditions by merely substituting a reel of film or a different set of slides. Similarly, one could provide for different kinds of responses to the same stimulus patterns. Such an adaptation of motion-picture tests might thus aid materially in the search for the important dimensions of motor skills by experimental methods. For example, a need in this regard is to hold the perceptual elements of a given range of motor tasks constant while varying the response dimensions. Many of the previous correlational analyses of motor task performances have employed tasks in which both the stimulating conditions and the responses required were allowed to vary.

Work Methods

The importance of work methods in success in fine motor skills has been stressed by Seashore (26, 28). In fact, it appears to be Seashore's belief that "hitting upon" favorable work methods may account for a greater proportion of individual differences in certain groups of operations than any combination of basic motor abilities. This hypothesis seems to raise a question about how well we can ever expect to predict success in certain motor tasks from any standard testing procedures. Although the work-methods hypothesis is important in relation to training in fine motor skills, it might be well to consider it from the point of view of selection.

One might test the hypothesis that for a given range of tasks, a person brings with him a general way of attacking such problems and that differential "work habits" may be predictive of success in complex psychomotor or manipulative tasks. The crucial problem here initially is that of describing and measuring such work methods. This is as yet

an unsolved problem. One might merely wish to consider efficient work methods identical with efficiency of learning. Thus, a standard motor task might be presented to a subject and the speed with which the task is learned to some criterion might be considered predictive of future complex motor learning. Or, again, one might merely wish to observe when the sudden progress associated with "insight" learning occurs during the task. Presumably this would be the stage when useless movements are dropped out and the most efficient work method is applied.

Another approach might utilize participant-observer techniques or introspective reports of individuals performing standard complex motor tasks. A classification of habitual approaches to a range of difficult psychomotor tasks might be organized and a weighted checklist constructed which could then be validated against job success.

Another approach might make use of automatic recording devices such as "operation recorders" which trace response patterns on a tape moving at a given rate of speed. Such a recorder hooked up in the proper fashion to a complex coordination test might indicate (a) what sequence of movements the subject used, (b) when the subject attempted to move simultaneously in more than one plane at a time, (c) when wrong directional moves were made, (d) when the subject was moving too far in a given direction, and (e) when the movements became smooth and coordinated.

Records indicating the subjects' work methods during performance of a well-controlled psychomotor task would be even more meaningful when the work methods of subjects who score high are compared with those of subjects who score low on total task performance.

Selection Versus Training with Respect to Motor Skills

Related to Seashore's "work-methods" hypothesis is the possibility that muscular coordination, manipulative dexterity, etc. are best considered as training problems rather than as selection problems. There is some feeling that, although people differ in their motor skills, the kinds of motor skills involved in certain trades and machine operations are highly amenable to training. After training, people may become relatively homogeneous with respect to these skills. Thus Tiffin (39) has indicated that the tradesman usually succeeds or fails in proportion to his training and general mechanical comprehension, not in proportion to his basic dexterity. This does not mean that successful tradesmen do not need skilled movements but rather that such muscular coordination as may be needed can be developed in training and it is a lack of some other ability (e.g., mechanical comprehension), rather than inability to develop the muscular aspects of the job, that may prevent proficiencies in the job. Similarly, Harrell (16) feels that success in such jobs as textile-working involves more the ability to visualize certain aspects of the job and only to a negligible extent muscular dexterity. There have been too few studies on final level of motor skill attainment to evaluate these hypotheses properly. Even in the Air Force program, where psychomotor tests proved to have considerable validity in predicting training criteria for pilots, navigators, and bombardiers, little study was undertaken to relate the tests to more ultimate criteria in advanced training or in

operational settings. Thus, it may still be that differences in these skills contribute less to the variance in job performance as training is continued.

Even though it may be possible to reduce individual differences appreciably through training, there is still the question of individual differences with respect to training time required. From a cost point of view, this is still a highly desirable criterion to predict by means of psychomotor tests. Moreover, individuals requiring more intensive training on motor skills in order to complete training in a specified time might be identified by such tests. Research is needed to investigate the relative importance of a training versus selection approach in a given range of motor skills. Experimental laboratory studies investigating individual-difference variables in final and intermediate levels of skill attainment as well as studies predicting job and training success would seem relevant research areas basic to the future of psychomotor test development.

SUMMARY

In this paper, an attempt has been made to examine the area of psychomotor skills research from the point of view of aptitude test development. The historical background of psychomotor test development is presented, problems attending the use of such tests are discussed, and previous factor analysis studies in this aptitude area are summarized. Suggested directions of future research are discussed in terms of possible research approaches and certain methodological problems having research implications.

While test development in other aptitude areas has been increasingly aimed at tests sampling one ability category at a time, tests of psychomotor skills have become more complex in the number of abilities sampled at one time and in the types of testing apparatus used. It appears that a basic need in this area is an extensive dimensional analysis of motor abilities which would reveal a more basic classification of factors primarily important in the performance of a wide variety of psychomotor tasks. Research of this nature has been fragmentary and for the most part has investigated only a limited range of psychomotor tasks. However, the review of previous factor analyses of psychomotor tests does reveal surprising agreement between studies on previous factors isolated. The previous factors identified, those needing confirming, and gaps in the limited range and complexity of tests utilized offer some starting points for a more extensive dimensional analysis of motor skills.

In addition, further sources of hypotheses about motor skill dimensions are suggested. These include on-the-job surveys, transfer studies and other kinds of experimental investigations of motor skill variables, research on qualitative characteristics of motor responses, and small-scale factor analysis studies.

With respect to methodological issues, the value of complex "job-analogy" type apparatus tests decreases when we are faced with a classification rather than a selection problem. The problem of classification intensifies the need to find the most independent set of motor ability categories, tests of which could be weighted differentially in predicting efficiently for a wide variety of jobs. Also questioned is the assumption that the abilities involved in performance of complex psychomotor tasks are qualitatively different from the abilities involved in factorially more

simple tasks. Research on this problem systematically relating performance on complex tasks to performance on a range of "pure" tasks is needed.

The possibilities of paper-and-pencil tests and the role of motion-picture presentation in the psychomotor area are also discussed.

Possible research on the quantification and measurement of "work-methods" is pointed out. The implication here is that for a given range of tasks individuals bring with them a general way of attacking such problems, and that differential "work habits" may be predictive of success in complex psychomotor or manipulative tasks. Such measurements of "work habits" might thus provide additional predictions of later performance.

Studies relative to the problem of selection versus training with respect to motor skills are suggested. Experimental studies involving individual differences on final and intermediate levels of skill attainment and studies predicting job and training success at different stages are examples of relevant research needed on this problem.

It is hoped that this paper will help provide a framework within which some future research on psychomotor skills testing can be carried out. Several lines of possible research have been discussed, the primary one being concerned with a dimensional analysis of motor skills as a basis for future test development. It is apparent that much additional research in this aptitude area needs to be done before the utility of such testing can be properly assessed.

REFERENCES

1. BILODEAU, E. A. Acquisition of skill on the Rudder Control Test with various forms of social competition. USAF Air Training Command, Human Resources Research Center *Research Note* P&MS 51-6, July 1951.

2. BILODEAU, INA McD., GOLDBECK, R. A., & REYNOLDS, J. B. An exploratory investigation of the effect of test administrator on subject's performance. USAF Air Training Command, Human Resources Research Center *Research Note* P&MS 51-5, July 1951.

3. BROWN, J. S., & JENKINS, W. O. An analysis of human motor abilities related to the design of equipment and a suggested program of research. In P. M. Fitts (Ed.), *Psychological research on equipment design*. Washington: U. S. Govt. Print. Off., 1947. (AAF Aviat. Psychol. Res. Rep. No. 19.)

4. BUXTON, C. E. The application of factorial methods to the study of motor abilities. *Psychometrika*, 1938, **3**, 85–93.

5. CAMPBELL, M. The "personal equation" in pursuit performances. *J. appl. Psychol.*, 1934, **18**, 785–792.

6. CATTELL, R. B. *Description and measurement of personality*. Yonkers: World Book Co., 1946.

7. CHAPMAN, R. C. The MacQuarrie Test for Mechanical Ability. *Psychometrika*, 1948, **13**, 175–179.

8. DUDEK, F. J. The dependence of factorial composition of aptitude tests upon population differences among pilot trainees. I. The isolation of factors. *Educ. psychol. Measmt*, 1948, **8**, 613–633.

9. FRENCH, J. W. The description of aptitude and achievement tests by means of rotated factors. *Psychometr. Monogr.*, No. 5. Chicago: Univer. of Chicago Press, 1951.

10. GIBSON, J. J. (Ed.) Motion picture testing and research. *AAF Aviat. Psychol. Prog. Res. Rep.* No. 7, 1947.

11. GOODMAN, C. H. The MacQuarrie Test for Mechanical Ability. II. Factor analysis. *J. appl. Psychol.*, 1947, **31**, 150–154.

12. GREENE, E. B. An analysis of random and systematic changes with practice. *Psychometrika*, 1943, **8**, 37–52.

13. GUILFORD, J. P., & LACEY, J. I. Printed classification tests. *AAF Aviat. Psychol. Prog. Res. Rep.* No. 5, 1947.

14. GUILFORD, J. P., & ZIMMERMAN, W. S. Some AAF findings concerning aptitude factors. *Occupations*, 1947, 26, 154–159.

15. HARRELL, T. W. A factor analysis of mechanical ability tests. *Psychometrika*, 1940, 5, 17–33.

16. HARRELL, T. W. Testing the abilities of textile workers. *St. Engng Exp. Stat. Bull. Georgia*, 1940, 2, No. 2.

17. HUMPHREYS, L. G., BUXTON, C. E., & TAYLOR, H. R. Steadiness and rifle marksmanship. *J. appl. Psychol.*, 1936, 20, 680–688.

18. MELTON, A. W. (Ed.) *Apparatus tests.* Washington, D. C.: U. S. Govt. Print. Off., 1947. (*AAF Aviat. Psychol., Prog. Research Report* No. 4.)

19. MICHAEL, W. B. Factor analyses of tests and criteria: a comparative study of two AAF pilot populations. *Psychol. Monogr.*, 1949, 63, No. 3 (Whole No. 298).

20. MUSCIO, B. Motor capacity with special reference to vocational guidance. *Brit. J. Psychol.*, 1922, 13, 157–184.

21. PERRIN, F. A. An experimental study of motor ability. *J. exp. Psychol.*, 1929, 4, 24–57.

22. REYMERT, M. L. The personal equation in motor capacities. *Scand. scient. Rev.*, 1923, 2, 177–194.

23. SEASHORE, H. G., & KOCH, G. Postural steadiness under conditions of muscular tension and fatigue. *Psychol. Rec.*, 1938, 2, 319–332.

24. SEASHORE, R. H. Stanford Motor Skills Unit. In W. R. Miles and D. Starch (Eds.), University of Iowa studies in psychology. *Psychol. Monogr.*, 1928, 39, No. 2 (Whole No. 178), 51–66.

25. SEASHORE, R. H. Individual differences in motor skills. *J. gen. Psychol.*, 1930, 3, 38–66.

26. SEASHORE, R. H. Work methods: an often neglected factor underlying individual differences. *Psychol. Rev.*, 1939, 46, 123–141.

27. SEASHORE, R. H. Experimental and theoretical analyses of fine motor skills. *Amer. J. Psychol.*, 1940, 53, 86–98.

28. SEASHORE, R. H. Work and motor performance. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1341–1362.

29. SEASHORE, R. H., & ADAMS, R. O. The measurement of steadiness: a new apparatus and results in marksmanship. *Science*, 1933, 78, 235.

30. SEASHORE, R. H., BUXTON, C. E., & MCCOLLOM, I. N. Multiple factorial analysis of fine motor skills. *Amer. J. Psychol.*, 1940, 53, 251–259.

31. SEASHORE, R. H., DUDEK, F. J., & HOLTZMAN, W. A factorial analysis of arm-hand precision tests. *J. appl. Psychol.*, 1949, 33, 579–584.

32. SEASHORE, R. H., STARMAN, R., KENDALL, W. E., & HELMICK, J. S. Group factors in simple and discriminative reaction times. *J. exp. Psychol.*, 1941, 29, 346–349.

33. SEASHORE, S. H. The aptitude hypothesis in motor skills. *J. exp. Psychol.*, 1931, 14, 555–561.

34. SEASHORE, S. H., & SEASHORE, R. H. Individual differences in simple auditory reaction times of hands, feet, and jaws. *J. exp. Psychol.*, 1941, 29, 342–345.

35. SPAETH, R. A., & DUNHAM, G. C. The correlation between motor control and rifle shooting. *Amer. J. Physiol.*, 1921, 56, 249–256.

36. SUPER, D. E. *Appraising vocational fitness.* New York: Harper, 1949.

37. THORNDIKE, R. L. *Personnel selection.* New York: Wiley, 1949.

38. THURSTONE, L. L. The perceptual factor. *Psychometrika*, 1938, 3, 1–17.

39. TIFFIN, J. *Industrial psychology.* New York: Prentice-Hall, 1947.

40. WALKER, R. Y., & ADAMS, R. D. Motor skills: the validity of serial motor tests for predicting typewriter proficiency. *J. gen. Psychol.*, 1934, 11, 173–186.

41. WITTENBORN, J. R. Mechanical ability, its nature and measurement; II. Manual dexterity. *Educ. psychol. Measml.*, 1945, 5, 395–409.

42. ZACHERT, VIRGINIA, & SHIBE, E. Comparison of the United States Employment Service General Aptitude Battery and the Airmen Classification Battery. USAF Air Training Command, Human Resources Research Center *Research Note* PERS 51-4, February 1951.

THE USE OF THE FREE OPERANT IN THE ANALYSIS OF BEHAVIOR[1,2]

CHARLES B. FERSTER[3]
Harvard University

The increasing number of experiments in the literature using the Skinner box makes it more important that the methods and techniques of this kind of experimentation be public. Since the technical problems of the experiments do not bear centrally on the analysis and would prove burdensome to the general reader, it is understandable that published accounts of this type of research do not include detailed specifications of all the techniques employed. Yet it is becoming impossible to repeat experiments solely on the basis of the published accounts. This paper will describe some of the criteria and principles which would make it possible for the interested researcher to duplicate the conditions of free operant experiments.

The Skinner box is generally considered a special technique employing an apparatus design of the kind originally used by Skinner (1). Any piece of apparatus differing in any way from the original design has been termed a modified Skinner box. Contrary to this view, the Skinner box is not a specific technique, but rather a method of research employing the free operant. The use of the free operant is a method of wide generality; it refers to any apparatus that generates a response which takes a short time to occur and leaves the animal in the same place ready to respond again. The free operant is used in experiments when the main dependent variable is the frequency of occurrence of behavior. Nearly all the problems of a science of behavior fit this paradigm when the questions are of the form: what is the likelihood of a piece of behavior occurring on this particular occasion; how strong is the tendency to behave on this occasion relative to another occasion? The free operant has advantages in this respect, because it removes restrictions on the frequency with which a response can occur and permits the observation of moment-to-moment changes in frequency (4, 5). The behavior of approaching the food magazine and eating, which is insensitive to a large number of variables, provides one such restriction. When the free operant is placed on an intermittent schedule of reinforcement, the small number of reinforcements relative to the number of responses minimizes the restriction on frequency. With proper techniques, animals can be trained so that the behavior of approaching the magazine and eating is under good stimulus control and, for all practical purposes, an invariant.

Since the form of the free operant is arbitrary, a further advantage can be gained by choosing a response which has a zero or near zero unconditioned rate. Frequency of response under an intermittent schedule of reinforcement generates a dependent variable which can take values over a very wide range, and which can serve as a base line for judging the effect of many variables. The range of the dependent variable is large. The frequency of a pigeon's pecking response, for example, can vary from zero or a few responses per hour to 20,000 responses per hour; over shorter periods rates have been observed as high as 15 responses per second.

INSTRUMENTATION OF THE FREE OPERANT SITUATION

Instrumentation of the free operant situation requires the following: a manipulandum, i.e., some device which the animal can manipulate and which operates a switch; a recording system; a magazine to present and control the reinforcing stimulus; a small chamber into which the animal and apparatus can be placed; and control equipment to arrange the required contingencies between the operation of the manipulandum switch, magazine, recorder, and any stimuli that might be presented. These components will be taken up one at a time. Since the principles of instrumentation will be clearer from examples, apparatus and procedures for the pigeon will be described. The methods are general, however, and the applications to other organisms may be induced from the pigeon techniques. In addition, an extensive series of research reports covering work on the pigeon will be published shortly which will illustrate the methods and techniques of this paper.

Manipulandum

The problem in the design of a manipulandum is one of generating responses by reinforcement that will be of similar form so that all of the members of the class will be recorded and covary with the manipulations of the independent variables. As an example of the kind of difficulty that can be encountered in this respect, consider the case of a pigeon key which requires a forceful peck to actuate the switch. Since the magazine is operated only when the peck is of sufficient force to operate the key switch, the force of the peck will be differentiated. Some variation will occur, however, with some pecks stronger than necessary and some pecks not having the required force to operate the switch. It must be decided whether the pecks which do not actuate the switch, and therefore go unrecorded, are to be counted as pecks. The problem is one of the generic nature of the response class and can be settled only by examining the lawfulness of the behavior that has been recorded. The difficulty can be avoided by designing the key so that the lightest peck that is expected will operate the switch.

Since automatic programming and objective recording of the experiment require that the response being measured operate some switch, other problems in the definition of a response can occur. It is possible, for example, to design a manipulandum that can be operated by behaviors which have extremely different topographies. In such a case it would not be clear whether unexpected observations were artifacts from variations in topography or attributable to their experimental manipulations.

The selection of the pecking response in the pigeon is a particularly fortunate one since it is already in

the repertory of the pigeon and occurs with little variation in the topography other than the force of the peck. The pecking response would, of course, be extremely unsuitable where variables influencing the form of the behavior were being manipulated.

The occurrence outside of the experiment of the behavior which is measured is another factor which must be controlled by training procedures and apparatus design. This is done by selecting a response which is infrequent in the natural repertory of the animal, by making the manipulandum prominent, or by placing the behavior under complete stimulus control by proper training procedure. The pecking response of the pigeon does not offer difficulty in this respect since the unconditioned level of pecking on a wall surface is almost zero; after the pigeon has been taught to peck, the light behind the key comes to control the behavior to the extent that if the color is changed the pecking stops.

The design of a key for a pigeon offers special problems because of the extremely high rate of response that can be generated. Rates of response as high as 15 per second can occur under appropriate schedules of reinforcement. The advantages that accrue from the wide range of values over which the rate of pecking varies are lost if the key does not have a sufficiently high natural frequency. A high natural frequency is obtained by making the armature of the key extremely light and the distance the armature must travel in order to close the actuating contact as small as possible.

Several types of keys have been in use in the Harvard laboratory, none of which are satisfactory in every respect. A key that is used at present employs a bakelite frame on which is mounted, on bearing surfaces, a piece of clouded Plexiglas. Relay contacts are mounted on the bakelite frame on phosphor bronze with the opposing contact on the Plexiglas. When the key is operated, the flexing of the phosphor bronze supplies a slight wiping action which cleans the contacts. A spring holds the Plexiglas armature against the bakelite frame and the resulting contact arrangement is normally closed. The maximum frequency of operation of the key can be increased by increasing the tension of the spring. Conversely, however, pecks will have to be executed with greater force in order to be recorded. The usual solution is a compromise between maximum frequency of operation and minimum force necessary to actuate the switch. Where it is possible to predict the rates of response that will be encountered, one practice is the use of a heavy spring on the key armature where uniformly high rates of response are expected, and a reduced spring tension where uniformly low notes are expected. Because of the millions of times the key will operate, the key contacts must be protected from heavy current loads and sparking. This is done by isolating the key from the rest of the programming circuit by a spark-suppressed relay of high enough impedance. Since the required duration of a peck before it can be recorded depends on the operate time of the keying relay, the relay chosen must be such that a minimum time elapses between the application of a voltage and the operation of the relay. The key switch is normally closed, because the time between the application of a voltage to the relay and the closure of its contact may be as much as 40 milliseconds longer than the time

between the discontinuation of the voltage and the closure of the normally closed contact. Since the speed of operation of a relay tends to be inversely related to its impedance, a compromise is made by choosing a relay whose impedance is low enough to ensure fast operation yet high enough that excessive current loads are kept from the key contacts.

A key which is adequate for many applications can be constructed by hinging a piece of Plexiglas at the top and placing a limit switch at the bottom, behind the Plexiglas. The movement of the Plexiglas either makes or breaks the limit switch contact. The characteristics of the key can be adjusted by changing the thickness of the spring or the distance between the contacts on the spring.

The optimal size of key has not been explored, but a key 1 inch in diameter has been found to be satisfactory. The pigeon reaches the key through a 1-inch aperture cut in the wall. The height of the key from the floor depends on the size of the bird. For the homing pigeon this distance would be $7\frac{1}{2}$ inches from the center of the key.

Recording

The most straightforward recording arrangement that could be used would be a polygraph. This would give a complete record from which any kind of computation or presentation could be made. The amount of work necessary to transcribe the polygraph record into a form that is usable makes this kind of recording unfeasible, however. This would be true for any experiment employing a free responding situation, but particularly true for experiments with pigeons, where as high as 20,000 responses an hour can be recorded. Polygraph records also result in a loss in efficiency because the experimenter can have little notion of the state of the experiment until a summary is made by laborious transcription. An even more serious difficulty is that the use of a polygraph precludes manipulations of variables in the middle of experimental periods.

It is the practice in the Harvard laboratory to devise recording instruments for each experiment which summarize the data in that aspect of the dependent variable which will be used in the analysis. The kind of recordings that are taken usually arises after considerable exploratory work. The record most frequently taken is cumulative. Other types have been used, however, and the following experiment is an example of use of a summary record of a noncumulative type.

The latency of a response to a discriminative stimulus was being studied. After preliminary exploration, the experiment was programmed so that the light behind the key came on once a minute. The response to the light was reinforced and the light was turned off. The time between the onset of the discriminative stimulus (the light behind the key) and the occurrence of the reinforced response was measured and an arbitrary value within this range was selected. This was called the criterion time. If the time between the onset of the light and the occurrence of the response was greater than this value, the response was not reinforced and the criterion time was lengthened by a fixed amount. If the response occurred within the criterion time, the response was reinforced and the criterion time was shortened. The data were analyzed by registering the changes in criterion time. The paper

was driven at a constant speed by a clock motor; the pen, which travels perpendicular to the direction of the paper drive, moved a constant distance up the paper whenever the criterion time was increased and moved a constant distance down the paper whenever the criterion time was decreased. The resulting record could be examined at a glance and showed concisely the change in latency over the experimental period.

Whenever an intermittent schedule of reinforcement is used, a cumulative record generally proves to be the most convenient and useful method of recording. The cumulative record represents nonprocessed data, in the sense that it is drawn by the bird directly. If a proper scale is chosen, it is possible to recover all of the information that would be at hand if a polygraph record were taken. In most cases, however, this choice of scale would vitiate the advantages of the record. The cumulative record is drawn as follows: A pen is stepped across the paper a small distance for each response. At the same time the paper feeds at a constant speed. The slope of the line that is drawn is directly proportional to the rate of the response. The virtue of the cumulative record is not that it allows a precise measurement of the rate at any particular time, but rather that it emphasizes changes in rate which can be seen in the curvatures of the record. Continuous rate changes occurring over a period as long as several hours can be summarized in the raw data of a cumulative record. In addition, subtle variations such as the "grain" of the rate show up on local curvatures.

The choice of a scale is dependent upon the range of rates that are expected. These in turn are dependent upon the species of animal used, the

operant chosen, and the schedule reinforcement. While the rate of sponse is directly proportional to tangent of the slope over the wh range, it is the angle that the tang makes with the abscissa that is ac ally observed in inspecting the l variations in the record. The rat change of the rate of response v respect to the angle of the tang to the curve increases very rap as the angle becomes large. At va higher than 80 degrees the tangen the angle increases so rapidly t changes in rate over short period time (the fine grain of the record) impossible to measure. The scale t is selected is therefore such that highest rate expected will produc line whose angle is no more than degrees with the abscissa.

A convenient scale for use v pigeons on interval schedules of r forcement would be a paper feed approximately 11 or 12 inches hour and 1,000 responses for a 6-i excursion of the pen. For use v rats, this scale would be much too duced and a better choice would either a slower paper feed or fe responses per excursion of the pa Where both fine grain effects over-all trend are desired, it is o the practice to use two recorders, of which gives a much redu record.

In order not to contaminate record with eating time, the gramming equipment is arranged that the paper drive stops during operation of the magazine. A ma is used on the recorder to indicate response which was reinforced.

In discrimination experiments the practice to use tandem record one for S^D and one for S^Δ. Both S^D and the S^Δ records are intern orderly and show a degree of law ness which would not be immedia

apparent if the behavior under the two stimuli had not been recorded separately. In experiments on concept formation four recorders have been used in tandem. The limiting factor in the most useful recording arrangement, in this type of experiment, is the expense of the recorder.

Magazine

The magazine is the device by which the reinforcement of the free operant is instrumented. In the design of the magazine it must be recognized that in conditioning a response a chain of responses is being formed, and that the critical event is the one immediately following the response. When we wish to reinforce a peck we must follow it by a prominent event which is the discriminative stimulus for the first member of a chain of responses leading finally to the ingestion of food. The concern with the chain of events, which maintains the reinforcing properties of the stimulus following the peck, is simply that it occur with regularity and uniformity under good stimulus control.

The types of magazine that can be used have two kinds of effects, a fixed amount of food or a fixed period of access to food. While the first of these guarantees that the same amount of food will be eaten for each reinforcement, it makes difficult the maintenance of good stimulus control over behavior in respect to the magazine. The amount of time the animal spends at the magazine will vary because there is no clear-cut stimulus correlated with the end of the food delivery. In addition, it is impossible to know whether the correct amount of time has been subtracted from the records. A more serious difficulty is encountered when the bird leaves food in the magazine

and finds it later, in the absence of the discriminative stimulus for approaching the magazine. The partial abolishment of the magazine discrimination results in behavior which will be in competition with the pecking response.

The magazine which presents food for a fixed period of time eliminates all the difficulties that are encountered in the fixed-amount presentation magazine, but suffers a disadvantage in that the amount of food ingested for each reinforcement is not necessarily uniform. Nevertheless, the fixed-time presentation magazine is undoubtedly superior for most purposes, because the variability introduced by slight variations in the amount of food per reinforcement is small in respect to the difficulties encountered from competing behavior conditioned in the same terms as the dependent variable of the experiment.

The natural diet for the pigeon is almost entirely grain, and grain turns out to be a suitable reinforcing stimulus, both from the point of view of the bird's behavior and the construction of a reliable magazine. The grain used in the Harvard laboratory consists of a mixture of 50 per cent kaffir corn, 40 per cent vetch, and 10 per cent hemp seed. The grains tend to be of uniform size.

The magazine now at the Harvard laboratory works as follows: A solenoid draws a pivoted tray into a horizontal position where the pigeon can reach the grain through a small aperture; a funnel feeds grain as needed at the rear of the tray; a 6-watt lamp, mounted directly over the aperture through which the pigeon eats, lights whenever the tray is raised to the eating position; when the tray is dropped out of reach the light goes out and the resulting il-

lumination is sufficiently dim so that the pigeon cannot see the grain. The illuminated magazine serves as an S^D for approach to the magazine and its termination serves as an S^Δ in respect to pecking in the magazine. After only minutes of training the bird makes no attempt to reach the grain when the tray is in the removed position and the magazine light is off.

Other types of magazines which move the grain in and out of the bird's compartment, or which cover or uncover the grain, have been used but suffer the disadvantage that the bird can be hurt by having its bill caught in the moving part of the magazine. While this will occur only rarely before the bird learns to avoid being caught by the moving part, the resulting timidity in respect to the magazine will make the magazine training much more difficult and introduce a variable into the experiment which can be easily eliminated by the design of the instrument.

The Experimental Chamber

The size of the chamber in which the bird works is to a large extent arbitrary, so long as the magazine is positioned in the neighborhood of the manipulandum. Limitations in space, however, usually require that the apparatus be kept as small in size as possible. A suitable size for a bird would be approximately $15 \times 12 \times 12$ inches. When the bird compartment is this size, artificial ventilation is almost always required.

Sound insulation is a knotty problem whose solution depends upon the taste of the experimenter and the requirements of the particular problem. The physics of sound absorption inevitably requires insulation by materials of considerable weight and thickness. This solution is both very expensive and space-consuming. For most applications the degree of sound proofing that is afforded by a well-constructed picnic icebox will provide data which are unaffected by random noises that occur in the experimental room. Critical events which can form the basis of a discrimination, however, must be silent. Since these events are nearly always the operation or the release of a relay, the problem can be solved most easily by using high impedance, light-weight relays that are shock mounted. The low intensity of these clicks, together with the amount of sound insulation applied by a good quality picnic icebox, will make impossible the formation of the discrimination based on a relay operation. In addition to these precautions it is the practice of the Harvard laboratory to supply in each experimental chamber a masking noise consisting of a random spectrum of sound. This type of noise is most effective in masking clicks. For nearly all applications, there is no need to eliminate the recorder noises. The click which the recorder makes every time a response is made serves as a conditioned reinforcer for the response and helps maintain a stable form of the response.

Control Equipment

Because the response that is chosen in the free operant situation actuates a switch every time it occurs, it is nearly always possible to instrument an experiment so that it runs automatically. Once the programming equipment for an experiment has been devised it requires very little attendance by a psychologist, and essentially unskilled personnel can run the experiment. Apparatus does double duty, since it can run all night without any attendance. Night experiments are often an efficient

solution where it is decided to minimize outside sound disturbances. The limiting factor in the number of experiments that can be carried out under conditions of automatic programming is the ingenuity of the psychologist, the amount of money available for programming devices, and competent technical assistants.

Night programming of experiments requires some method of terminating the experiment in the absence of the experimenter. The problem is easily handled in the case of the pigeon because the pigeon roosts in the dark. To terminate the experiment, it is necessary only to disconnect the reinforcing circuit and turn out all of the lights in the apparatus. When a nocturnal animal is used, however, external events are correlated with the termination of the experiment and a discrimination is formed.

There are no simple rules or procedures for the automatic programming of experiments. The program is accomplished almost exclusively by the relay and timing devices, and there is no substitute for general relay know-how or technical assistance. Facility for designing relay circuits for use in programming experiments comes with a small amount of practice.[4]

Care must be taken in the design of relay circuits so that the length of pulse necessary for the operation of the relays in the recording circuit is not larger than the length of pulse necessary for the operation of the relays in the reinforcing circuit. If the operate time of the reinforcing circuit is shorter than the recording circuit, the force or duration of the

pecks will be differentiated so that large numbers of pecks will occur which are of sufficient force to operate the magazine, but not capable of operating the recorder. One solution of this problem is to require a larger duration and intensity of pecks for the reinforcing circuit than is required for the recording circuit. In this way the mean force or duration of pecks that occur is increased, and the likelihood of nonrecorded pecks is correspondingly decreased.

The Pigeon as an Experimental Animal

The pigeon has advantages over the rat as an experimental animal that justify the use of a new animal in spite of a considerable body of literature already dealing with the rat.

The pigeon lives as long as 15 years. This makes possible long-term experiments, into which the maturation of the rat would introduce radical changes. A variable interval schedule, for example, which took 30 or 40 days to reach a steady state, would result in an experiment which would encompass an important fraction of the life span of the rat. Considerable economy is achieved by using pigeons sequentially in experiments. For example, if an experiment on some aspects of schedules of reinforcement resulted finally in stable behavior which can serve as a base line, the birds could be reused in an entirely different experiment which required a base line of that sort.

The pigeon has very good visual acuity and color vision. This leaves great latitude and variety in the kinds of stimuli that can be used in discrimination experiments, and the number of different stimuli that can be used within a single experiment. In contrast, the rat's very poor

---

[4] An excellent text on general principles of relay circuit design is available in *The Design of Switching Circuits*, by W. Keister, A. E. Ritchie, and S. H. Washburn, D. Van Nostrand, 1951.

vision offers serious difficulty in the design of experiments analyzing complex processes.

The pigeon comes to the experimenter with a well-tailored response extremely suitable for free operant type experimentation. The high rates of pecking that can be generated in the pigeon result in a dependent variable which can change over a very wide range, and which is necessarily more sensitive to manipulation.

THE PREPARATION OF BIRDS FOR AN EXPERIMENT

The birds should be banded as soon as they are received from the supplier and their wings clipped along the second run of feathers so that they cannot fly. They are fed freely for three or four days, after which time the ad lib. weight is taken. This is designated as the normal weight. After this the birds are not fed again until they reach approximately 80 per cent of their ad lib. weight. This will take about one week if no food is given. An alternative practice, however, is to feed 5 grams of grain every other day. While this will make the deprivation period longer, the disruption of the bird's digestive system will not be so severe. The degree to which any particular animal's weight must be reduced can be ascertained only by attending to behavior. A given percentage of reduction in body weight will not have the same behavioral effect on each animal. Therefore, the weight should be reduced until the animal will eat out of the magazine without hesitation. If the bird does not eat out of the open magazine within 15 minutes after being placed in the apparatus, it should be returned to the loft and not fed for another day.

After the bird eats freely and has had its daily ration on two separate days (15 grams each day, for the typical bird), the magazine training proper can be started. The task at hand at this point is to establish the magazine light and sound as the occasion on which approach to the magazine will occur with high probability, and to extinguish responses to the magazine in the absence of the magazine light and sound. To accomplish this the magazine is operated approximately every 30 seconds. The bird will approach the magazine in the absence of the light, and this behavior will soon extinguish. At this stage of its training the bird should be watched carefully to avoid the development of any superstitious behavior (3). If any superstitious behavior develops, it can be eliminated simply by withholding the magazine operation until some other behavior is in evidence. When the bird moves rapidly toward the magazine as soon as it is operated, seldom approaches it when it is not operated, and eats continuously throughout the operation, the magazine training is complete and the pecking response can be differentiated.

The pecking response can be differentiated by any one of three methods. In all three methods the key is connected so that its operation presents the magazine via a switching circuit. In this way, when the first peck occurs the discriminative stimulus for approaching the magazine will be immediately contingent on the response.

1. The first method depends on the unconditioned level of responding. The bird is simply left in the apparatus with the key connected to the magazine. The first key peck that occurs will almost result in con-

ditioning if the magazine training has been effective.

2. The second method uses a discrimination already in the bird's repertory. A small grain is scotch-taped to the key. The bird pecks at the grain and incidentally operates the magazine. Where removal of the grain disturbs the bird, it can be "vanished" by replacing it with progressively smaller grains.

3. The third method is the differentiation of the response. This is done by observing the bird and selecting some behavior which approximates the peck. Often the first step is opening the magazine when the bird is facing the key. The procedure from this point is one of alternate conditioning and extinction of various responses, which progressively approximate key pecking. As soon as the bird is facing the key regularly, reinforcements are withheld until some variation occurs in the behavior which is in the direction of the required response. If reinforcements are withheld too long and the bird extinguishes, the process is begun again.

It is very important that the operation of the magazine be immediately contingent on the specified behavior. If the reaction time of the experimenter is too long, the magazine operation may strengthen other behavior than was intended.

With a small amount of practice and a hungry, well magazine-trained bird, it should take no more than five minutes to "shape up" a pecking response. If there is difficulty in shaping up the pecking response, the bird is probably not hungry enough. The hunger of the bird cannot be judged from the fact that the bird eats readily out of the open magazine, since this behavior is relatively insensitive to changes in the level of deprivation. The frequency of occurrence of the kind of response that is being conditioned is a much more sensitive indicant, and the percentage of body weight reduction that is finally assigned to the bird is best determined by looking at the general amount of activity. If the bird is inactive, conditioning will be difficult and the bird should either be reduced in weight or habituated to the apparatus. A suitable level of deprivation is more easily determined after an intermittent schedule of reinforcement is established. The subjects can be equated in terms of level of deprivation by adjusting their body weight so that the rates of response under a variable interval schedule of reinforcement are equal.

The initial conditioning may be carried out either in the cage that will be used in the experiment or in a specially prepared conditioning apparatus which affords a good view of the bird. Which one is used depends on the conditioning procedure and the type of experimental cage. No particular advantage is gained by the use of an open cage if the conditioning is carried out by the "unconditioned rate" or "grain of corn" method. If the differentiation procedure is used, however, an open cage is necessary, unless the experimental cage has a window which will allow a clear view of the bird when the cage is at a level as high as or higher than the experimenter. The birds will be seriously disturbed if the experimenter looks down on them.

If the pecking response is conditioned by differentiation, some difficulty may result from the accidental conditioning of unnatural pecking topographies or superstitious behavior. This is not likely to be the case where the conditioning is accomplished by the use of a kernel of

grain or the unconditioned rate of pecking. Where an unusual topography does occur, several sessions of continuous reinforcement in the experimental cage will almost always result in a uniform, natural topography.

The length of an experimental session is limited only by the number of reinforcements that are given. The total amount of food that is given is such that the bird's weight will be the same the next day. If the bird does not receive enough food during the experiment to bring its weight up to the value determined as the normal weight, the difference is fed at the end of the session. If the bird gains weight, the number of reinforcements given each session must be reduced. About 60 reinforcements will usually maintain the weight of a bird if a 3-second presentation of grain defines a reinforcement.

THE DESIGN OF EXPERIMENTS

There are no rules for the design of experiments on the free operant other than those of accepted scientific practice. Intermittent reinforcement of the free operant, however, allows many experiments to be designed around base lines generated by a schedule of reinforcement. Two examples will clarify how the intermittent reinforcement of a response supplies a base line for judging the alteration of behavior due to an experimental manipulation.

Amount of Reinforcement—the effect of variations in the size of a reinforcement on the tendency to behave.

The pigeon is trained on a variable-interval schedule of reinforcement which generates a constant rate of response. An arithmetic-interval schedule of reinforcement produces a constant rate of response after approximately 15 hours of training. In the middle of an experimental period the amount of reinforcement is either decreased or increased. In this way the effect of the independent variable on the behavior is uncontaminated by any passage of time or handling of the animal. The experiment is then continued until the behavior is stable under the new procedure, and the reversibility of the process is checked by returning to the original amount of reinforcement. If it is found that the process is not reversible, the procedure of alternating the amount of reinforcement is continued in some counterbalanced order. This allows an analysis of the dynamic effect of changes in the amount of food in each reinforcement. If the behavior under different amounts of reinforcement is reversible, the procedure is repeated a sufficient number of times to establish its significance. The number of reversals that are needed depends upon the size of the effect and the lawfulness of the change in the behavior from one amount of reinforcement to another. The parameter is explored over the complete range in a single animal; and the experiment is repeated upon a sufficient number of animals to convince the experimenter that the effects of the manipulation are large in respect to the variation due to individual differences. Relatively few animals are necessary if the records show that the dependent variable changes in the same way from subject to subject, and differs only in terms of a constant, such as over-all level of responding. Sufficient control is possible for use of groups as small as two animals, although a group of three is more efficient statistically. Further refinement of the experiment is necessary if the subjects differ in the dependent

variable in both the constant effect and the nature of the change.

Discrimination

The development of a discrimination between two colors will serve as another example of the use of an intermittent schedule of reinforcement as a base line. After the pecking response is differentiated, the bird is given several sessions of continuous reinforcement and the color of the light changes after each reinforcement. Reinforcement then occurs equally in the presence of the two colors on a variable-interval schedule of reinforcement, and it is not predictable whether the color will remain the same or change after each reinforcement. Separate recorders operating in tandem record the behavior under the two colors. When the rate of response under the two colors is the same, reinforcements are discontinued in the presence of one of the colors. The intervals of the presentation of the color, however,

remain defined by the previous reinforcement schedule. As the rate of response in the presence of the nonreinforced color (S^Δ) declines, the rate under the reinforced color (S^D) does not change. It provides a base line for judging the effects of variables such as level of deprivation.

Despite the alternation of the reinforced and nonreinforced stimuli, the decline in rate in S^Δ is extremely orderly. Moreover, after two or three hours of training the rate under the nonreinforced color will be less than 1/100 the rate under the reinforced color. Such an analysis of discrimination problems permits the experimenter to determine over very short intervals of time the tendencies to behave in the presence of both stimuli. The use of an intermittent schedule of reinforcement as a base line for the analysis of the development of stimulus control is easily extended to complex processes, such as concept formation, with similar advantages.

REFERENCES

1. SKINNER, B. F. On the conditions of elicitation of certain eating reflexes. *Proc. nat. Acad. Sci.*, 1930, 16, 433–438.
2. SKINNER, B. F. *The behavior of organisms.* New York: D. Appleton-Century, 1938.
3. SKINNER, B. F. "Superstition" in the pigeon. *J. exp. Psychol.*, 1948, 38, 168–172.
4. SKINNER, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, 57, 193–216.
5. SKINNER, B. F. Some contributions of an experimental analysis of behavior to psychology as a whole. *Amer. Psychologist,* 1953, 8, 69–78.

EXPERIMENTAL STUDIES OF SMALL GROUPS[1]

MARY E. ROSEBOROUGH

Harvard University

In the past few years the study of small groups as an area of experimentation has been attracting the interest of an increasing number of social scientists. A body of literature has been accumulating, upon which it is the purpose of this paper to report.[2]

The field of small-group analysis is here defined as the study of persons in face-to-face relationships. A group must be small enough to provide the possibility of intragroup communication, so that each member, after a group meeting, may report impressions and thoughts about every other member if so required. Many of the studies involve the direct observation of artificially formed groups in the process of solving problems presented by an experimenter under controlled laboratory conditions. It makes no difference for definitional purposes whether members are acquainted prior to the meeting of the group, but naturally this factor is of prime concern in experimental design.

Priority for inclusion in this review of experimental studies has been given to controlled laboratory inves-

tigations rather than descriptive field reports of groups in which there has been little attempt to define precisely the variables, both dependent and independent, or to seek relationships between two or more observed behavioral indices. Since the major purpose of this review is to evaluate the present state of empirical knowledge in the field of small group study, those articles in which emphasis is primarily theoretical or interpretive have been eliminated. Thus, the work of early thinkers such as Cooley, Elliot, Follett, LeBon, Mead, Simmel, Sumner, etc., who are important for their many shrewd insights and hypotheses about groups, is not dealt with directly but only indirectly in so far as it has stimulated and proven to be amenable to experimental investigation. Similarly, recent systematic treatments of small group phenomena such as those given by Bales, Bion, Homans, Moreno, Newcomb, Redl, Whyte, etc., and review articles, for example those by Gibb, Platts, and Miller (57), Shils (143, pp. 40–52), and Swanson (157), have little place in a paper of this kind.

Another criterion influential in determining the selection of articles is implicit in the definition of the field of small-group analysis previously stated. That is, for definitional reasons, the work of F. H. Allport, Sengupta, Sinha, Weston, English, Farnsworth, etc. is not considered here. Their work falls under the general heading of "social facilitation" since they were particularly concerned with the way in which the

mere physical presence of others acts as a stimulant on individual activity, eliminating for the most part interpersonal communication. For similar reasons many studies of the effect of spectators or auditors, and of rivalry and competition in human and animal subjects are excluded. Thus, studies which provide the possibility of intragroup communication form the core of the field of small-group research.

The studies of small groups here reviewed are classified in terms of the *independent variables specified* for investigation. The general plan of this survey is to subsume the investigations under five topical headings. These headings are as follows: (*a*) studies designed to contrast and compare behavior of groups and individuals, (*b*) studies which involve the manipulation of social structure variables important to group functioning such as authority relationships, (*c*) studies of the effect of cultural variables, that is, the sharing of values and goals in a group, (*d*) studies involving the manipulation of situational conditions such as group task, size of group, communication networks, etc., and (*e*) studies of personality variables affecting group behavior.

Behavior of Groups and Individuals

Many experiments have been designed to determine whether persons working cooperatively in groups are more efficient at problem-solving than persons working without interaction or in competitive, individually-oriented circumstances. The dependent variable in the typical experiment under this heading is a productivity index, such as number of mathematical problems solved or accuracy of judgment. This was the

main concern of the early experimentalists, but interest has waned with present emphasis on processes in small groups.

Efficacy of Group Performance versus Individual Performance

Problems with specified criteria of evaluation. Group discussion seems to have an effect on individual behavior in a wide range of activities and this effect is generally evaluated as desirable by the experimenter. Münsterberg (120) found that individual judgments of the numbers of dots on cards were more correct after group discussion than before. Group performance has also been found to be more efficacious than individual performance for word-building out of letters by Watson (163); for numbers of arithmetic problems correctly solved by Shaw (141), Barton (10), and Klugman (90); for memory and word construction by Radu (132); for accuracy of judgment of events by Klugman (91); and for accuracy of judgment of events in a legal situation by Dashiell (37).

All of these laboratory studies deal with a simple type of problem for which there are definite criteria to evaluate performance. However, the same results have been obtained using more complex situations. For example, Lorenz (102) studied the effect of the work group on productivity in a shoe factory and found an increase in output efficiency of a worker in a cooperating group over his output when isolated.

Granting that group membership has a powerful effect upon individuals, still the conditions under which such an empirical generalization holds must be explicitly stated. For example, Burtt (28) repeated Münsterberg's experiment and found little confirmation for the results previ-

ously drawn. Such contradictory findings can only be resolved by carefully designed experiments in which variables and conditions are explicitly and clearly specified.

Marston (110), criticizing Dashiell's experiment for lack of specification, showed that a single trained individual may be a more successful finder of facts than a number of untrained jurors, this superiority varying with the specific kind of fact finding in which he has been trained. Also, the knowledge an individual may have of how others' opinions in his group differ from his own is a condition which greatly influences the effective role of argument and persuasion according to Jenness (83). His study was designed to qualify the conclusion drawn by Münsterberg that "individual judgments are more correct after group discussion than before." Münsterberg had always arranged it so that the subjects in his experiment knew exactly what the others' judgments were before group discussion began.

Other authors, Gordon (59, 60) and Stroop (155), point out that in many of the studies in this area, for example Watson's study, the results would have been little altered if each person had worked separately and their products combined statistically. That is, when subjects made independent judgments of weights and these were artificially combined into statistical aggregates, the composite judgments increased in correctness with increased size of the aggregate. Gordon's criticisms refer to tasks which merely require the addition of individual contributions. Thus, "group" thinking is superior to that of an individual just by the fact that the thinking of a number of individuals is pooled. Under such conditions, as was pointed out by Marston, a

particular individual working alone may be superior to a group and would be impeded through group discussion with inferior persons. However, when a task is used which involves a number of logical steps, all of which must be in a correct order to find the right answer, such as in Shaw's study, group discussion may have real advantages over individual endeavor. The advantages seem to be due to the rejection of incorrect suggestions, the checking of errors in the group, and the increment in the range of suggestions and original ideas.

Other versions of the same type of study may be found. Deutsch (39) essentially induced an attitude of mind in his investigation of cooperation and competition similar to that created in the foregoing studies. The students of the cooperative groups received final grades that depended upon the solutions to puzzle and human relations problems which their group as a group submitted. In the competitive groups the final grade was determined by how much each individual contributed to the solution. The findings indicated that greater productivity occurs when the members of a group are organized in terms of cooperative activities rather than competitive ones. Evidence supporting this conclusion has been reported by Mintz (116), Stendler, Damrin, and Haines (151), and Maller (105). Maller concluded that cooperation with an organized team resulted in greater efficiency than work for the self, while cooperation with an arbitrary group, chosen by the examiner, resulted in lower efficiency. Efficiency was measured by the speed with which simple additions were done by school children.

Problems with unspecified criteria of evaluation. Group discussion has been shown to have an important

influence on individual behavior even where performance can be judged, only arbitrarily, as more correct, faster, or more effective. Bechterew and de Lange (18) gave subjects moral conflict situations to decide upon and found that after group discussion individual solutions were more inclusive of the many relevant facts in the case as well as more similar to each other.

Sherif (142), using the autokinetic phenomenon, had subjects, while working alone, make successive judgments of the extent of the apparent movement of a light. Under these conditions each subject developed a range within which he made his estimates. These subjects, when subsequently put in groups of two and three, gradually came to make their judgments within a restricted range, characteristic of the group. When, however, subjects began by making judgments together, they kept their group norms while later making judgments alone. Sherif uses this experiment as the prototype by which group norms, attitudes, and values are established. Propagation of similar type experiments has occurred (24, 64, 140).

These latter studies, although not posed in efficiency terms as the former experiments, are likewise designed to prove that a group situation, vaguely defined, does indeed affect individual performance. Such studies characteristically minimize the group discussion process itself, the analysis of which is the core of small-group research at present.

Efficacy of the Lecture Method versus the Group Discussion Method

Essentially the same question is being investigated in the work aimed at proving that group discussion has more powerful effects upon individual behavior than the traditional lecture method. In the actual teaching of a college subject by the lecture and class discussion methods, Jones (87) found that students scored higher on immediate and delayed tests for reproduction in the discussion groups. This result was substantiated only for delayed recall by Bane (9), however, and Spence (148), working with graduate students, found just the reverse result to that of Jones. Recently Husband (79) and Asch (3) have reported that the lecture method is more efficient in teaching than the method of small section meetings when final grades are used as the measure of achievement. However, Asch drew just the opposite conclusion, using measures of change in social and emotional adjustment as criteria.

Many of these contradictory conclusions are no doubt the result of poorly specified and poorly controlled experimental variables as well as the lack of uniformity in the operational meaning of "efficiency." The conditions under which group discussions are found to be superior to lecture methods ought to be explicitly stated.

Other experiments have been conducted in a variety of situations. The results are unequivocal in the studies of food habits by Lewin (95), Guthe (65), Willerman (166), and Radke and Klisurich (131). More housewives changed their behavior and attitudes about various types of foods after participating in a group discussion than after hearing lectures concerning these foods. In the lecture situation individuals resisted anything that might have made them depart from old group standards which they had internalized. After group discussion, however, change was facilitated since new group standards had evolved. Thus the

resistance which was due to the relation between the individual and the group standard was eliminated.

In addition, group discussion has been found to be effective in changing a wide range of behavior patterns. Prejudicial attitudes have been changed (99, 109), hostile attitudes lost (20, 86), community problems solved (80, 98), alcoholics have been cured (5), neurotic disabilities alleviated (22, 89), emotionally disturbed children helped (21), productivity raised (35, 81), roles and status changed (125), frustration induced most successfully (97), and disabilities accepted (41). We need not be further persuaded that group discussion processes have an effect on individual performance even though there is a selective process occurring in the reporting of studies. This proof has only opened up new and troublesome problems concerning the mechanisms by which this influence is achieved and the conditions under which such an empirical observation holds.

In the food-habit studies, for example, there were certain conditions which, needless to say, determined the results. In these discussion groups there was no question who the discussion leader was, and he functioned in that role from beginning to end. Furthermore, there was no question about the objective of the group discussion. It was designed to change the food habits of the members. These are but two of the possible factors which conditioned the results.

Many of the other studies lose considerable persuasive powers *by failure to make use of a control group design.* Also, the experimental variable, vaguely defined as "group discussion," has included groups with leaders or without, groups of various sizes, meeting over different periods of time, some with opportunity for feedback and self-evaluation, some permissive and others directive, groups with different purposes and with all types of participants, from children to neurotics. Finally, the results of these studies have not been as cumulative as might be expected since the problems set for investigation tend to be structured in a polemical way.

SOCIAL STRUCTURE VARIABLES

In this section, experiments are reported which attempt to make explicit the effects of one set of variables, variables which were uncontrolled and for the most part unrecognized in the previous section. More specifically, different types of authority relationships external to the group are investigated in terms of the effects these have on group productivity, shared attitudes in a group, therapeutic gains, group integration, amount of communication, etc.

Diffuse Authority Relationships: Studies in Attitude Change

In the process of interaction, group members develop norms, attitudes, and motives which they hold in common, according to Whyte (165). These norms may or may not be in accord with the norms of persons in authority positions. The following studies indicate the consequences on small-group behavior of two status levels holding concordant or divergent norms, goals, and interests about a common subject.

The pioneer studies in industry by Mayo (111), Roethlisberger and Dickson (135), and Whitehead (164) of two small groups of factory workers indicated that when strong loyalties exist between workers and manage-

ment, productivity is high but when negative sentiments prevail, the workers develop an elaborate informal system of rules and sanctions such that output is actually restricted. Here loyalties toward management were established by the workers by virtue of their selection for special participation in the studies, and of management's policy not only to consult these special workers before any change was introduced into the work routine but also to allow the workers to participate in the designing of the new job. Negative sentiments were engendered as the result of management's disregard of the workers in policy making.

Similar results have been reported to point up the advantages of co-operative action on the part of management with workers (**2**; **14**, p. **25**; **51**, **52**, **108**, **109**). In addition, resistance of piecework employees to changes in their work methods and jobs prescribed by management was found by Coch and French (**35**) to be inversely related to the amount of joint decision-making granted to workers by management. With active participation, the workers developed task goals among themselves and with management and overcame their resentment against authority so that successful job change was possible. Maier (**104**) showed further that a leader skilled in conference procedure and possessing exceptional ability in solving technical problems could conduct a discussion so as to obtain not only group acceptance of his ideas but also to obtain a quality of problem-solving that surpassed that of a group working with a less skilled leader and/or without technical competence.

This result has been found not only in industrial settings. McCandless (**113**), working within a training school setting with high-grade, mentally deficient children and predelinquents, wanted to decrease the social acceptability of the domineering type of leader in the informal peer groups. He found that the relationship between dominance and popularity was high in cottages directed by adult supervisors, but that this relationship decreased as the boys themselves were allowed to participate in supervision and policy making. Preston and Heintz (**130**) found that "participatory" leadership as opposed to "supervisory" leadership was more effective in changing attitudes toward presidential nominees, in producing group agreement, and in making the task more interesting among college students.

Specific Authority Relationships: Studies in Leadership Style

The studies reported in the previous subsection were phrased in terms of changing attitudes of group members in order to affect some other desirable goal. The following studies have to do with leadership style which can be manipulated to induce different types of specific authority relationships in order to study its effects on other small-group behavior.

The very famous Lewin, Lippitt, and White (**96, 100, 101**) study of autocratic, democratic and laissez-faire leadership is an excellent example of this type of experimental manipulation. The autocratic leader determined all policies, techniques, and activities. He maintained his autonomy by remaining aloof from active group participation except when demonstrating to the group members what they were to do. In the democratically led groups all policies were determined by group discussion with the leader taking an active role. In the laissez-faire

groups, the leader took no active part, the group having complete freedom for group or individual decision.

The leaders, by virtue of their own behavior toward the groups, induced response patterns not dissimilar to the results obtained with industrial workers. The autocratic leaders provoked hostile behavior among the boys (cf. Mayo and Lombard [112] in the next section) or apathetic behavior (cf. Mayo [111]). The laissez-faire leaders created interpersonal friendliness among the group members but dissatisfaction with the task was high for these groups. The democratic leaders were preferred in a popularity rating. Their groups were task oriented, well satisfied with their achievements, and highly integrated. In terms of quantitative work output, the submissive autocracies surpassed the other groups. However, the products of the democratic clubs are reported to have been qualitatively best because of "the close attention given to every detail."

Nevertheless, the proselytizing of a faith in democratic leadership is not the scientific aim of small-group study as here defined. The scientific aim is rather to delimit the conditions—social, cultural, psychological, and situational—under which this empirical relationship holds. For example, would the results be confirmed in an authoritarian organization like an army? Only with careful definition will results contribute to the cumulative growth of small-group study.

Thelen and Withall (159) have made a contribution to this aim by investigating the differences in perceptions by group members under various styles of leadership and consequently the extent to which leaders can indeed induce desired experi-

mental conditions. They found that gross structural characterizations of leadership styles by persons possessing very different bases for evaluation (these bases were labelled objective-behavioral, projective-attitudinal, subjective-internal) were in substantial agreement. Heyns (72) observed the effects on participant behavior of discussion leaders who manifested an active and positive relationship to their groups as contrasted with leaders who had essentially negative attitudes. The most notable result was the greater incidence of supportive and cumulative contributions among members in the positively led groups and of opposing contributions in negatively led groups. Also there was greater acceptance of opposing behavior in the negatively led groups such that opposers received high popularity ratings and were the most highly satisfied with the decision.

Other studies of leadership style in small discussion groups have been conducted in an educational setting with the purpose of improving teaching methods. The results at present seem to be inconsistent. Wispé (167) used two styles of teaching in discussion groups at the college level. He found no differences in final examination marks between students in directively led groups and students in permissively led groups. However, students preferred the former. On the other hand, Allport (1) reported results to indicate that students preferred informal, permissive, and friendly discussion groups, as did Gross (62), who also found significant gains in insight on a self-insight scale administered before and after the group meetings. Evans (42) concluded that directive leadership in therapy groups is more efficacious than nondirective leadership, using

therapeutic gains, member preference, and attendance records for the basis of this conclusion. On the other hand, Faw (43) found that the non-directive approach was preferred by students, that it induced high student participation and better performance on examinations. Bovard (25, 26) showed that affectivity was higher between members in "group-centered" groups and members were more susceptible to opinion changes than in "leader-centered" groups. The advantages of "group-centered" groups in school learning have been pointed out by Flanders (46) and Perkins (126). Rehage (133), however, obtained results which did not warrant claims of superiority for one method over another.

Put in *efficiency* terms these results are confusing. (The same contradictory conclusions were encountered in the polemical studies of the efficacy of the lecture method versus the group discussion method.) Like many evaluative studies, the conditions under which advantages are greater for one type of leadership style than for another are for the most part poorly stated. The experimental results of these studies stand as particular findings, not unified into a conceptual scheme of problems and concepts of social relationships, and incapable of such unification because there is lack of systematic study of the effects of independent variables.

Recent empirical and theoretical writing (6, 7) about small-group behavior has tended to regard the small group as a dynamic system of action, which ebbs and flows between instrumental-adaptive activity as task problems are being solved, and expressive-integrative activity as socioemotional problems among the members are attended to. That is, over and above the common norms and attitudes which are built up in the group, there are equilibrium properties of the small group itself such that imbalances of activity in certain areas have repercussions in other areas of activity.

Evidence for such a view has been presented by Sterling and Rosenthal (152). They found that leaders and followers change with different phases of group process, the same leader usually coming to the fore as similar phases of the group recur. Bales and Strodtbeck (7) present evidence to support the notion of phases in group problem-solving. Under certain conditions there are subperiods within the total period, in which interaction changes its character in predictable ways as a group proceeds from initiation to completion of a continuous period of interaction. A similar view is held by Heyns (71) and Plank (127).

With such a conceptual scheme, perhaps questions posed in dichotomous terms, such as directive or permissive leadership styles, are not formulated most realistically. The behavior style of a leader may be appropriate for one phase in the sequence of events but maladaptive for another. If a leader rigidly maintains one behavior pattern and is not sensitized to his functions in the various phases of the group, then something less than maximum group satisfaction or efficiency will conceivably ensue.

A similar criticism can be levelled at many studies of sociometric patterning and composition of groups, in which the assumption is made over and over again, for example by Zeleny (169), that compatible groups are the most productive or efficient groups. By maximizing the expressive-integrative dimensions in a small group it is assumed that this will *ipso*

facto facilitate the solution of instrumental-adaptive problems. If workers are given the opportunity to express their needs, their productivity rates will go up. This type of static, one-sided thinking is not consistent with the conceptualization of a small group as a dynamic system of action and has not been supported by empirical evidence (136).

In two remaining studies by Thibaut (160) and Kelley (88), the experimenter, rather than acting in accordance with a specific leadership style, differentially meted out rewards to young boys working in teams. In both, a team could receive one of four experimental treatments: it could consistently receive reward throughout the experimental period; it could receive consistently no reward; it could receive reward at the beginning of the experimental period but by the end be receiving no reward; and finally, it could gain in reward from the beginning to the end of the experiment.

Thibaut formed groups in such a way that at the outset each group had approximately half of its sociometric choices in its own group and the remaining half in an opposing team. He studied the effect of the experimenter's manipulation of rewards upon the proportion of own-group choices. Only with consistent reward or no reward, or what Thibaut called high and low status, did the sociometric attractiveness of one's own team increase. He also found that the amount of communication initiated by a group tended to increase as the group's position changed toward a lower status and tended to decrease as the team's position became more rewarding. The hypothesis presented was that people who are motivated toward upward mobility will tend to communicate

with those above them in the organizational structure.

Kelley concluded that the most disruptive circumstance to intragroup friendliness was what he called high status (reward) combined with the possibility of demotion (withdrawal of reward) and low status combined with the impossibility of promotion. The provision of status security for the highs or the addition of the possibility of moving upwards in the case of the lows led in his groups to the maintenance of group solidarity. He also found that the very existence of status differentials may operate to prevent ease of communication.

This concludes the section on types of authority relationships and their consequences on small-group behavior. One of the dependent variables in these studies is the presence of common attitudes, norms, and values among group members. In the next section, the degree to which a culture is shared in a group becomes an independent variable which is manipulated in order to study its effect on other aspects of small-group behavior such as problem-solving, amount of communication, content of discussion, reaction to frustration, etc.

CULTURAL VARIABLES

The cultural component in small-group experiments has been manipulated in two different ways. First, group affiliation has been used as an indication of shared values and norms on the assumption that these are formed and held in common as group members interact with one another. Group affiliation is a difficult variable to use, however, since the degree and type of acquaintanceship are often hard to specify. Consequently, a second method has been devised. Groups are formed and in-

structions given in such a way as to establish group frames of reference desired by the experimenter.

Unfortunately, the strengths of one method of variation are the weaknesses of the other. The second type of variation has advantages in that the extent to which norms and goals are shared can be precisely specified. It has the disadvantage of creating experimental situations in which group affiliations are artificial, goals limited, obligations weak, and permanency lacking. The reverse is true of the first type of variation. Thus, the study of the effect of a common culture has inherent difficulties in an experimental setting.

Group Affiliation

There are a number of studies which investigate the effect of prior acquaintanceship and therefore of shared values on other small-group behavior. In a study by French (49, 50) the organized groups were composed of teammates from a neighborhood house while the unorganized groups contained subjects who were unknown to each other before the experiment. Both groups were exposed to experimental conditions of frustration, produced by the attempt to solve objectively insoluble problems, and to conditions of fear produced by suffusing their locked room with smoke in simulation of a fire. The results indicate that the members of the organized groups compared with the unorganized groups were more interdependent, had greater we-feeling, had greater equality of participation among members, had greater motivation to complete the problems and greater feeling of frustration when unable to do so, were more inclined

to direct their aggression toward other persons in their group, and showed more fear. Using preschool children, Wright (168) studied the social behavior of pairs of strong friends and pairs of weak friends when subjected to frustrating situations. The results indicate that strong friends showed more cooperation and more aggressiveness than weak friends.

Using two behavioral indices—absenteeism and turnover—Mayo and Lombard (112) investigated differences in group solidarity between two large industrial plants. The factories were similar in geographical location, technology, and labor force but differed in the degree of primary group organization on the worker level. In the low-turnover factory, personal bonds, shared norms, constitutive of a primary group, were well established and newly recruited workers were easily incorporated. In the high-turnover factory, these bonds were weak and no sanctions were enforced as workers performed their functions independently of one another. This was especially true of the new workers who showed a very large proportion of absence and turnover. This result has been substantiated by Fox and Scott (48) and Davis (38) among others.

Strodtbeck (154) studied husband-wife dyads in three cultures as they attempted to resolve opinion differences. The cultures differed in the way the status of women was defined, women having a powerful position in Navaho society compared with men, less powerful in the Texan community, and least powerful in the Mormon group. He found that the differential ability of husbands and wives to win decisions and to participate verbally in the experimental

sessions was closely related to the definition of the power positions of men and women in the larger social and cultural organizations. He pointed out the further relationship between the actual amount of influence exerted by a person on a group decision and his pattern of participation, a result substantiated by Staton (149). The most influential spouse tended more frequently to ask questions, carry out opinions and analysis, and make rewarding remarks while the less influential spouse showed acts of agreement, disagreement, and aggressive acts designed to deflate the other's status. The wide range of possibilities that may be used to influence the behavior of another has been described by Merrill (115), who observed mothers and children in a standard play situation.

Finally, Gyr (66) attempted to obtain descriptions of the cultural differences in customary committee procedures by interviewing representatives from China, South America, the Near East, and the United States. The findings are extremely tentative due to the inadequacy of the sample. The greatest differences are evident in the attitudes of South Americans and North Americans. There appeared an awareness of leader superiority, trustfulness in delegating authority, and desire to cooperate in the United States sample, while in the South American sample there was general uncertainty about the motives of others.

Experimental Instruction and Arrangement

There are several studies in which groups were formed and instructions presented in such a way as to create low or high group affiliation or cohesiveness. "Cohesiveness" is used in these studies as an explanatory concept.

Back (4) used differences in the salience of instructions to establish low and high cohesive groups. It was found that in the highly cohesive groups, members not only tried harder to reach agreement and to influence their partners on controversial issues but were somewhat more willing to accept their partners' opinions than were members of the low cohesive groups. Schachter, Ellertson, McBride, and Gregory (139) created low and high cohesiveness within groups in a similar way in order to determine what effect cohesiveness had on standards of production. There was no indication of any necessary relationship between cohesiveness and productivity. However, attempts by the group to influence a member to decrease her rate of production was more effective in high than in low cohesive groups. Cohesiveness appeared not to be a determining variable when the request was made to increase output.

Festinger and Thibaut (45) manipulated the solidarity dimension by the force with which the experimenter emphasized in the instructions the necessity, first of all, for a task solution and, second, for a unanimous group decision. The results indicated that the volume of communication directed toward a group member was a function of his holding an extreme opinion. The greater the solidarity, the greater was the tendency to communicate to persons expressing extreme opinions and the greater was the actual change toward uniformity of opinion in the group as a whole.

Finally, Schachter (138) created his degrees of cohesiveness by manipulating the attractiveness of the

activities which the groups mediated. That is, students becoming members of clubs of their choice constituted high cohesive groups while those becoming members of clubs which they were not interested in joining made up the low cohesive groups. Using stooges to express deviant opinions, he showed that in high cohesive groups, the deviant was more strongly rejected on sociometric testing than he was in low cohesive groups. In terms of amount of communication addressed to the various group members, he found that more communication was addressed to the deviant in highly cohesive groups, and that this trend was more pronounced during the discussion of issues vital for the group.

Thus, group affiliation and the sharing of norms and goals by members seem to have behavioral correlates. Group affiliation and its concomitants were defined independently in the former studies, while in the latter studies (with the exception of Schachter) cohesiveness was inferentially derived. These latter studies failed to create many of the characteristics of primary group membership. They tended in that direction to the extent that the group members were influenced in their attitudes in the way predicted by the experimenter; that is, to the extent that the group members accepted the definition of the situation and of the goals pronounced by the experimenter. The experimental conditions seemed to achieve an approximation which went far in effectively changing participant behavior in the expected directions such that it appeared justifiable to include these studies in this section. The circularity in this argument is apparent, and a closer examination of the concept of solidarity or cohesiveness created

by experimental manipulation is required.

SITUATIONAL VARIABLES

In this section, studies are included which are designed to define more precisely the influence of variables which are part of the physical situation impinging on a group. Four such variables have been studied: (a) the nature of the task presented to the group, (b) the numerical size of the group, (c) the spatial arrangement of members, and (d) the external restrictions on communication channels.

Task Problem

Every conclusion made about small-group behavior depends upon the instrumental or task problems confronting the group under study. Generally the task problem itself is not the independent variable but is a situational condition which is included in the interpretation of results, or is disregarded and unspecified.

There are several experiments actually designed to determine the nature of task influences on small group behavior. Carter and Nixon (31) used three different tasks—intellectual, clerical, and mechanical-assembly—as situational variations which were shown to affect leadership behavior. In general, the subjects who took the lead and influenced their partners in the intellectual task also influenced their partners in the clerical tasks. The mechanical-assembly task gave only low inter-correlations with the other two problem-solving situations. In a later study Carter, Haythorn, and Howell (29), also interested in the relationship between leadership and group task, employed factorial analysis, a type of analysis used in small-group

study, particularly by Cattell (33) in his search for general group and leadership dimensions. Two major factors were revealed—an intellectual leadership and a leadership based upon manual skills. The conclusion to these studies was that leadership is specific to the situation. However, as pointed out by Gibb (56), this does not exclude some degree of generality of leadership in similar or related situations.

Deutsch (40) found that the differences in task structure between mathematical problems and human relations problems determined to a great extent the quality of group process. In the solution of the mathematical problems, there was more individualized effort, less coordination of efforts among group members, fewer attempts at communication, and more communication difficulties. The content of the human relations problems was more "value-laden," which provoked more conflict in the group, that is, blocking, self-defending, and aggression among the members. The previously cited study by Shaw (141) likewise revealed major differences in group process contingent on the task presented for solution.

Heise and Miller (67) used three kinds of tasks: (a) a simple reassembling of a list of standard words, (b) construction of a sentence, the words of which had been distributed among the group members, and (c) anagram formation. Using controlled communication channels between members (see the subsection on Communication Patterns) they concluded that the relative efficiency of a communication pattern depended upon the kind of problem the group was trying to solve. While the network pattern of communication was an important condition in the solu-

tion of the reassembly and construction problems, it had little effect on anagram formation. The reassembly task was most efficiently solved in groups where all members could talk and listen to all other members. The construction problem, however, was solved most efficiently in a group which had a man in a central coordinating position.

Finally, Bales and Strodtbeck (7) found that the sequence of events in problem-solving groups varied in character with the type of problem under consideration. The tasks they analyzed ranged in substance from group projection sketches of Henry and Guetzkow (70) and chess, to group decision and planning problems with various degrees of reality.

Thus the problem presented to the group has been shown to be a pervasive situational condition which affects small-group behavior, variously measured by, for example, leadership, efficiency, and sequential scores. So far, few studies have investigated the task problem as an independent variable. What results are available sound a cautionary note for all small-group experimenters to specify clearly the problem orientation of the groups under investigation and interpret results accordingly.

Size of Group

Very few experiments have actually been designed to discover the differences attributable to change in group size, despite the very early intuitive analysis by Simmel (144) of the effect of group size on the sociological form of the group. Moede (118), using the physical pulling power of men as his measure of effectiveness, found that a four-man group was most economical. He claimed that the pulling power of the average individual of a group de-

creased by 10 per cent with each additional worker. Marriott's (107) study of a total of 330 various sized management-organized working groups in two motor-car factories demonstrated an inverse relationship between output per man and group size.

However, as was pointed out by South (147), the efficiency of various sized groups depends upon, among other things, the type of problem the group is attempting to solve. Comparing three-person and six-person groups, he concluded that the smaller groups are more valuable when the problem lends itself to immediate formation of solution while the large groups perform more efficiently when the problem requires that wrong hypotheses be promptly rejected. The other factor he specifically investigated as affecting efficiency was the sex composition of groups. He concluded that mixed groups are less efficient. However no systematic treatment has been given to this variable even though most experimenters make implicit assumptions about its effect by purposely studying either all male or all female groups.

More recent investigations have not been framed in productivity terms but in terms of the effect of size on social structure. Hemphill (69) concluded that within the limits imposed by his methodology, the size of the group is a variable which to some degree conditions leader behavior. As the group becomes larger, demands upon the leader's role become greater and more numerous, and tolerance for leader-centered direction of group activities becomes greater. Bass and Norton (12) concluded that the relative stratification in a group, measured by the mean and variance of leadership ratings made by observers of participants,

tended to increase with increases in discussion group size. Essentially the same result has been found by Bales, Strodtbeck, Mills, and Roseborough (8), who have reported that the difference in total amount of participation between the most talkative person and the least talkative person in a group increases as the group size increases from three to ten men. There is need for further specification of the effect of this important situational condition on the behavior of small groups and greater sensitivity to its influences.

Spatial Position

Working on the assumption that a person will be more likely to interact with another if he is in a good position to see what the other is doing as well as hear him, Steinzor (150) investigated the spatial factor in face-to-face discussion groups. He found that the degree of interaction among the members sitting more nearly opposite from one another in a circle differs significantly from that expected by chance. As an ecological factor, spatial position is not unrelated to networks of communication, to be discussed in the next section, or to propinquity as a determiner of friendship patterns in the community at large (44).

Communication Patterns

A situational variable being investigated by a team of researchers headed by Bavelas (15, 16, 17) is the communication network. The standard situation used is one in which the communication paths between members of the group are controlled so that the effect of predetermined patterns of communication can be studied. Communication is usually in written form. The standard task presented to the group consists of the

simple collection of information, each member contributing some components to the correct solution. The task as a conditional factor in the interpretation of the results has been studied by Heise and Miller (67). (See the subsection on Task Problem.)

Leavitt (92), using four communication patterns—circle, chain, Y, and wheel formations—found that the circle and wheel patterns produced the most contrasting results. Using, as measures of effect, records of speed, errors, and number of message units in the solution of the problem, as well as postmeeting questionnaires, it was found that the circle pattern produced the most active, erratic, unorganized, leaderless but satisfied group. On the other hand, the wheel pattern was least erratic, required relatively few messages to solve the task, was organized with a definite leader, but was less satisfying to most of its members. The member in the central position in the communication network became the leader and he was more satisfied with his job than were the occupants of peripheral positions. The various communication networks did not, however, differ significantly in the average time taken to solve the problem. Similar results were obtained by Smith (146, p. 197), who used the same communication patterns. In addition he found that the circle pattern permitted the members to adapt more readily to a change which required the relearning of certain parts of the task and which upset a previously established learning set.

Heise and Miller (67), using networks of telephone channels, varied the intelligibility of messages by manipulating the relative intensities of speech and of extraneous noise. They found that by exposing a group to such unfavorable intelligibility conditions the differences in the relative efficiency of different networks can be exaggerated.

Again using the Bavelas technique, Leavitt and Mueller (93) showed that a two-way communication circuit between the sender of information and the receiver or executor of information, a condition which they called "feedback," increased the accuracy with which information was transmitted as well as increased the confidence of the receiver and sender in what they had accomplished. Although the condition of no feedback was less time consuming, it engendered hostility in the receiver and doubt in the sender.

The experimental situation investigated by this group of workers is considerably more restricted than most group situations studied in the laboratory. The differences between communication networks are maximized while it is assumed that other antecedent conditions—social structural, cultural, personality, and situational—are for the most part held constant. The development of a common culture between members is minimized by the nature of the standard situation and by the nature of the task which involves merely the collection of information. Any conclusions drawn from these studies must be in terms of the carefully controlled conditions under which the groups must work. This is a powerful experimental method if precision is not achieved at the expense of too little applicability.

Communication patterns have also been studied in free discussion situations by Bales, Strodtbeck, Mills, and Roseborough (8). Certain average empirical tendencies have been found. For example, he who initiates most action will receive most and will

address more of his remarks to the group as a whole than to specific individuals, etc. Therefore, even though the communication pattern in a free discussion group is uncontrolled at the beginning, there is a tendency for a pattern to be formed in the course of the group meeting such that certain channels have a higher probability of use than others. Thus the results about communication patterns derived from this treatment and the paradigm of Bavelas *et al.* seem capable of generalization.

PERSONALITY VARIABLES

The investigations included under the heading of "Personality Variables" are more often than not correlational studies which relate personality factors such as attitudes, needs, abilities, dominance, masculinity, talkativeness, etc. to measures of leadership in a group, interaction patterns, clique membership, and social status. The aim has been to study the internal dynamics of group behavior or some of the endogenous elements in the group deriving from the personal characteristics of the members and their roles within the group.

While previous investigations have studied the consequences of independent variables on the group as a whole or on problem solving in general, these studies focus rather on more microscopic effects. Personal material about members is used both for independent variation and for the measurement of effect. This double-barrelled interest in personality factors has tended to foster the disregard of the other factors which influence behavior in groups, such as social structure, culture, and situation. This neglect of other conditions in the design of experiments has led to faulty predictions and to overgeneralization.

The majority of the studies have to do with leadership behavior within groups. The previous studies of leadership were classified under the social structure heading since these concerned authority positions external to the small group.

Studies of Leadership Behavior

Interest in the problem of leadership has been keen for years. Since there are excellent reviews of the literature available (68, 82, 153) and since the main concern here is with small-group research, a great bulk of the material will not be mentioned.

One of the oldest polemics in this area of study is whether personality qualities are the pervasive factors in the determination of leadership, qualities which some people have and other people have not, or whether leadership is situationally determined. There is much evidence for the former view. Bell and French (19) attempted to determine the extent to which individuals maintain consistent leadership status in a series of informal discussion groups made up of different members. At the end of each session, the members of each group were asked to nominate a discussion leader for a hypothetical second meeting. On the basis of these results, the writers concluded that leadership status was highly consistent despite the situational changes involved. Of course this finding is relative to the other conditions operating in this study such as the size of groups, the type of task, permanency of group, population sample, etc., as has been pointed out before. However, the same consistency between leadership as rated in one situation and as rated in another was found by the OSS staff (123) in assessing officers during World War II. There were seven

social situations used for assessment: an interview, a problem-solving situation, a construction problem situation, panel discussion, debate, assigned leadership situation, and ratings by associates.

Merei (114) studied the extent to which children with leadership ability, so designated by nursery school teachers and observers in this study, had to change their behavior when placed in groups of children who had formed traditions and rituals. The results indicate that the leader adopted the new group's traditions rather than instilling his own. While in this sense he proved to be weaker than the group, he still managed to play the role of leader.

If leadership is consistent, what are some of the qualities that leaders possess? Attempts at specifying leadership qualities have been made by Chapple (34). He has devised a method for selecting supervisory leaders which has been employed for a number of years with apparent success in business organizations. Chapple assumes that an individual has a constant relationship between the frequency and duration of his verbal actions and inactions regardless of the situation. He records these interaction units in an interview on a polygraph which he calls an interaction chronograph. By comparing the pattern of interaction of a prospective supervisor with the interaction pattern developed to define the requirements of a particular supervisory position, he selects and places candidates. Chapple is inconsistent in his argument to the extent that he recognizes that different situations require different patterns of interaction, but he still tries to generalize from a two-person interview situation to all others. His assumption about the constancy of individual participation patterns loses force in the face of other evidence.

Using a method of recording similar to Chapple's, Horsfall and Arensberg (77) observed leadership patterns in teams of industrial workers. Although teams were structured around a highly interactive person there was no uniform rate of interaction within a team over and above the finding that the more productive teams had a smaller amount of "within" team interaction. Leaders showed a combination of a high interaction rate between and within teams, rather than a high rate of "initiating responses" (as distinguished from "terminating responses") or a particularly high overall interaction rate.

There are other investigations which have been designed to delimit and define leadership qualities. A correlational study with young children in nursery school groups was reported by Goodenough (58). She found correlations of the order of 0.60 among five variables—physical activity, talkativeness, laughter, social participation, and leadership. Bass (11) had judges rate participants in unsupervised discussions on a series of 13 items designed to reveal leadership status. These ratings were correlated with the time each participant spent in talking in the group session. Coefficients ranged from 0.82 to 0.92. In another study, Bass and White (13) had fraternity members judge each other on leadership items after which they participated in free group discussions. The ratings made by trained observers of the participants were found to be highly related to the judgments of members. Using similar experimental conditions for his study, French (53) reported confirmatory results. Green (61)

pointed out the further relationship between effective group participation, measured by duration of participant's contributions, and verbal intelligence measured by standard verbal intelligence tests, while Brown (27) reported high significant correlations between total participation and high manic scores on the Minnesota Multiphasic Test.

Hemphill (68) used questionnaires in studying leadership qualities. Respondents acted as observers of groups of which they were members. Each respondent gave a description of his group and reported his observations of the leader's behavior. The behavior required of all leaders regardless of the situation in which they functioned had the following characteristics: (a) ability to advance the purpose of the group, (b) competence in administrative functions, (c) ability to inspire the members of a group to greater activity or to set the pace of the group, (d) behavior which added to the individual member's feeling of security in his place in the group, and (e) behavior relatively free from activities serving self-interest.

Finally, in a field study of children in summer camp, Polansky, Lippitt, and Redl (129) found a positive relationship between leadership in a group (sociometrically determined) and the extent to which a person will directly attempt to influence the conduct of others and in turn show spontaneous reactions to the behavior of others. These data stimulated a further laboratory study by Grosser, Polansky, and Lippitt (63) of behavioral contagion and the qualities important in the acquisition of group prestige. They found that a stooge was able to influence the behavior of another just by his own acts of initiation.

A number of studies have already been presented, however, which demonstrate the importance of situational factors in the determination of leadership behavior (8, 31, 29, 69, 92). In addition, Hollingworth (74), in a field study of children's groups, found that children with high intelligence were seldom leaders in groups of average children but were likely to be leaders in groups of superior children. Carter and Nixon (32) related leadership ratings by school teachers and fellow students to observations of the same subjects participating in small leaderless groups. They found that power-seeking, money-oriented, persuasive, masculine people were often rejected as leaders by supervisors and associates, while in leaderless group situations they became leaders. This provoked a further study by Carter, Haythorn, Shriver, and Lanzetta (30) of the differences between leaders who are appointed in and those who emerge from a free discussion group. Although having the same amount of leadership ability, the leaders who emerged were found to be more authoritarian than leaders who were appointed.

Thus personality factors are of prime importance as conditions influencing the behavior of leaders or potential leaders. However, they are but one set of a complex of conditions which interact with one another to determine the behavior which is eventually manifested, the other sets being social structural, cultural, and situational factors. Consequently, any attempts to make predictions from personality qualities, without other reference, to leadership behavior in small systems of social action, have a high probability of failure (117). Research which clearly specifies the interaction of factors

holds great promise for small-group study. Several such investigations concerning leadership behavior are available. One of these is a study by Maas (103).

Maas was primarily interested in the interrelationships between personality and situational variables affecting the perception and behavior of group leaders. The personality of leaders was described in terms of need structure. Leaders with needs to change others rather than self he classified as projective, and those with needs to change self rather than others as introjective. He used autobiographical material and the California Personality Test as the basis for this classification. The groups varied in structure, the informal groups having open membership with activity programs, and the formal groups having closed membership with social issue programs. He found that projective leaders perceived the behavior of others and acted toward others in a causal and objective way in the informal groups where their needs could be satisfied, but in an evaluative and judgmental way in the formal groups where their needs were not met. The introjective leaders behaved in just the opposite way. That is, they were biased in the informal groups in their perceptions and actions and were objective in the formal groups.

Swanson (156) likewise was clearly aware of the interaction of personality and situational factors in the behavior not necessarily of leaders, but of members belonging to discussion groups. The Blacky Test provided personality data, and observations of the group prior to and during the experiment were used as situational data. The groups differed in degree of internal heterogeneity of social status and of ideology about group functioning as well as in the permissiveness of intermember behavior. Given the predispositions reflected in a particular Blacky Test profile, and granted the kind of interpersonal situation that a particular group provided, Swanson was able to predict significantly such measures as a person's volume of participation in the group, the amount of received disagreements from others, the amount of positive affect towards others expressed on sociometric tests, the judgment of others' task contributions, and the degree to which self was perceived as being influenced by others.

The behavior which has been predominantly studied so far has been generally referred to as "leadership behavior." However, the concept of leadership has had little uniformity in operational meaning. It has had such diverse definitions as "nomination as a discussion leader for a hypothetical second meeting," "designation by nursery school teachers," "those in supervisory positions," "ratings by judges in unsupervised discussions," "ratings by fraternity members," "sociometric status," "appointed leaders," etc. The lack of correspondence between such definitions has been demonstrated by Polansky, Lippitt, and Redl (128). Also the problems asked about leadership behavior have been posed generally in polemical terms. Consequently these studies do not provide, for the most part, a substantial, well-tested body of supporting evidence, despite the great wealth of experiments. Future investigators may well be advised to seek relationships between operational measures of the behavior of group members without recourse to a poorly specified, nontechnical term such as leadership.

Studies of Social Interaction Patterns

There are ever-increasing numbers of studies which attempt to relate in a systematic way, phenomenological or projective personality data, with interpersonal reference, and social interaction patterns. These are mainly correlational studies in design, little attempt being made at either manipulation or prediction.

A very early experiment by Riddle (134) shows the subtle interaction of individual motivation, evaluation, and betting behavior in a series of poker games. Riddle reached the conclusion that the "desire to win" in this game situation was only to a slight extent aroused by the size of the player's own hand value. It was aroused more fully by the value of the opponent's hand. The influence of one's own hand was to increase one's own bet, and the effect of the opponent's hand was to inhibit one's bet. When the balance had been struck and the bet made, this bet in turn duly determined the total strength of the "desire to win."

Coffey, Freedman, Leary, and Ossorio (36) and Ossorio and Leary (124) found relationships between the social interaction patterns of members in a group therapy situation and their personal ideologies obtained from test and autobiographical data. Changes at one level of personality structure were accompanied by changes at the other as the therapeutic sessions progressed.

Fouriezos, Hutt, and Guetzkow (47) reported that ratings of self-oriented needs from clinical material (Rorschach, TAT, sentence completion, and interviews) and from observations of persons in discussion groups were highly correlated. The self-oriented needs used by these investigators were dependency, status, dominance, aggression, and ca-

tharsis. They also found that self-oriented needs correlated negatively with group satisfaction and positively with the amount of group conflict. Malone (106) observed group psychotherapy sessions over a 12-month period. Therapeutic success, measured in terms of individual improvement, showed marked ups and downs during the observational period. He reported that when success prevailed "group sentiment, commonality of interests and objectives" were strong in the group; that success varied with the degree of group solidarity and the rate of social interaction. Finally, Joel and Shapiro (85) found a high relationship between patterns of interaction in a therapy session and patterns of interaction of characters in the MAPS Test and TAT stories.

The study by Horowitz, Lyons, and Perlmutter (76) indicated that agreement or disagreement with acts in the group is significantly related to one's attitude toward the person perceived as the source of the acts and the person to whom the acts are directed. These attitudes are built up on the basis of agreements and disagreements displayed by the person in response to one's own acts as well as the responses of another to the person's acts.

Again, Norfleet (121) related personal liking for a group member, judgment of productive contribution, and actual contribution to a discussion. She found that personal liking, measured by sociometric questionnaire, was poorly correlated with the actual number of contributions made by a person. However, judgments of productive contribution became more highly related to the total number of actual contributions as the period of study progressed. Working from another direction, Rosenthal

and Cofer (137) discovered that the nonparticipative behavior on the part of one group member produced measurable effects on attitudes toward the group goal of other group members.

Thus social behavior, values, and motivation are interrelated in a complex whole both within and between people. These in turn affect group properties such as group solidarity, commonality of interests, group satisfaction, etc. The nature of the interrelationships and the mechanisms of influence remain to be worked out.

Studies of Sociometric Choice Patterns

In a variety of the small-group studies already reviewed, sociometric techniques, as devised by Moreno (119), have been used as a means of revealing social structure. The sociometric studies to be discussed in this section make the choice pattern itself a focus for investigation. These studies are mainly correlational, direct relationships being sought between sociometric choice patterns and variables of personality. Sociometric measures have been used in at least two different ways, as status indices and as indices of clique membership. Sociometric status was found to be positively related to job satisfaction in groups of carpenters and bricklayers by Van Zelst (162) and to skill by Whyte (165), who used bowling teams in his study. Fulton (55) obtained correlations of the order of .54 between teammate status, measured by student choices, and skill in volleyball. However, such simple positive correlations were not obtained by French and Zander (52), who related popularity of individuals to their production level in industry. In some cases the correlations were positive while in others negative. They concluded, as does Homans (75), that status within a group varies more directly with the extent to which a person conforms to production standards set by the group than to the production level itself. Thus by linking these correlational studies more closely to what is known about social interaction processes, better sense can be made of the results.

Hunt and Solomon (78) observed a summer camp group of young boys. They found significant correlations between group status and previous experience in camp, athletic ability, generosity, physical attractiveness, orderliness of activity, and lack of egocentricity. With time in camp the correlation between group status and such a palpable characteristic as athletic ability decreased while those correlations between group status and behavioral traits increased. Lemann (94) related sociometric overchoosing to personality traits and found that girls with high status were rated as generous, enthusiastic, and affectionate, while girls with low status were rated as stingy, apathetic, and cold. Tagiuri (158), using preparatory school boys, related several dimensions of personality to sociometric choice data. He found that the perception of others' preference for oneself is more accurate if one chooses the other than if one does not. Also a person belonging to a religious group exhibits preference for others who belong to the same group. Finally, there were significant differences revealed in the choice patterns of students rated by the school psychiatrist as adjusted and those rated as maladjusted. This relationship between social status and adjustment has also been found by Northway and Wigdor (122) as well as by French (54).

Thus, the choice of another in a social situation is affected by the

personality characteristics of the other in its overt expression. However, the choice of another is also affected by the personality characteristics of the person choosing and by the properties of the group situation, a conclusion drawn from a study by Maas (103) and from Jennings' (84) extensive field study in a girls' training school. A theoretical approach which can deal with such a complementarity of factors seems better warranted than the unilinear approaches so often used in experimental design and explanation.

Friendship patterns have also been investigated by the use of sociometric techniques. The general procedure in these studies has been first to identify cliques and then to look for common characteristics among the members. Smith (145) analyzed such factors as sex, residence, athletic and nonathletic activities, grades in school, church preference, and father's occupational status as possible variables influencing the choice of friends among high school students. He concluded that students select friends who are in some ways like themselves and that consequently friendship selection might be merely a form of "ego-expansion." However, Bonney (23) found little relationship between the same factors in the mutual friendships of elementary school, secondary school, and college students, while Thompson and Nishimura (161) concluded on the basis of their study that friendships are formed, not so much because two friends are alike in terms of manifest traits, but rather because each approximates the ideal of the other.

These studies of friendship determinants are similar both in methodology and conflicting results to the studies of marriage mate selection.

For example, Hollingshead (73) identified two polar theories about the selection of marriage mates. One he called the theory of homogamy that "like attracts like" and the other he called the theory of heterogamy that "opposites attract each other." His data support the theory of homogamy, having used race, age, religion, ethnic origin, and class as his variables. By using marriage mates, a greater degree of control impinges on the data than is the case with the sociometric clique studies. However, since these investigations are tangential to the main body of small-group studies, there seems to be no necessity for further discussion of results here. Nevertheless, a generalized theory of the precise influences operating in the selection of friends or marriage mates remains to be formulated. Perhaps by linking this research more closely with evidence and interpretation concerning primary group affiliations, authority relations, group process dynamics, situational influences, and personality expression, the results will become more cumulative than they have been in the past.

SUMMARY

The study of small groups is flourishing. Promising research techniques are available and others are being devised. These have been used to investigate a variety of substantive problems. The substantive problem areas in small-group study, defined in terms of the independent variables specified in experiments, are classified under five topical headings.

1. The studies which aim to contrast and compare the behavior of groups and individuals show that group discussion processes have an effect on individual behavior in a

wide range of activities. However, this demonstration raises new and troublesome problems concerning the mechanisms by which this influence is achieved and the conditions under which it operates.

2. Under the social structure heading, studies indicate that authority relationships external to the group have important consequences on small-group behavior measured by productivity scores, attitude ratings, quantity of communication, etc. The small group is regarded as a dynamic system of action in the recent experiments which have ingeniously established and changed authority relationships in laboratory settings.

3. Cultural variables, studied within organized groups with a past history of sharing norms or within artificial groups with attitudes induced by experimental arrangement, have been shown to have effects on other small-group behavior such as expression of aggression, absenteeism and turnover, decision-winning, amount of communication, etc.

4. So far, knowledge is meager about the influence of situational variables, such as group task, size of the group, spatial arrangement of members, and communication networks, on group behavior. Such knowledge is vital to every small-group study since the results are conditional on these factors. Too often situational variables have been un-

controlled and unspecified in the design of experiments.

5. The majority of the studies included under the personality-variable heading have to do with leadership behavior. The study of leadership as a social structure variable or a personality resultant has provoked the greatest interest in small-group research. This may be due to many factors, two of which may be a dominant value emphasis in our society or an adaptation to practical demands from the army, business, and industrial organizations. However, the results have not been as substantial as might be expected. Advances in knowledge are anticipated from studies which seek relationships between specific measures of behavior of group members and personality data.

The artificial division into structural components of society, culture, situation, and personality is not to suggest independent function of these variables. The structural-functional design is an exigency in social science research where simultaneous manipulation of all variables is impractical at the present stage of development. Rather, the theoretical approach which seems best warranted in the face of the current evidence is one which views the small group as a dynamic system of action, action determined by a complex of interdependent or interacting factors.

REFERENCES

1. ALLPORT, G. W. How shall we evaluate teaching in the social sciences. In Bernice Cronkhite (Ed.), *Handbook for young college teachers.* Cambridge, Mass.: Harvard Univer. Press, 1950. Pp. 36–56.

2. ANDERSON, K. A Detroit case study in the group talking technique. *Personnel J.,* 1948, **27,** 93–98.

3. ASCH, M. J. Nondirective teaching in

psychology: an experimental study. *Psychol. Monogr.,* 1951, **65,** No. 4 (Whole No. 321).

4. BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.,* 1951, **46,** 9–23.

5. BALES, R. F. Social therapy for a social disorder—compulsive drinking. *J. soc. Issues,* 1945, **1,** 14–22.

6. BALES, R. F. *Interaction process analy-*

sis; a method for the study of small groups. Cambridge, Mass.: Addison-Wesley Press, 1950.

7. BALES, R. F., & STRODTBECK, F. L. Phases in group problem-solving. *J. abnorm. soc. Psychol.*, 1951, **46**, 485–495.

8. BALES, R. F., STRODTBECK, F. L., MILLS, T. M., & ROSEBOROUGH, MARY E. Channels of communication in small groups. *Amer. sociol. Rev.*, 1951, **16**, 461–468.

9. BANE, C. L. The lecture versus the class discussion method of college teaching. *Sch. & Soc.*, 1925, **21**, 300–302.

10. BARTON, W. A., JR. The effect of group activity and individual effort in developing ability to solve problems in first year algebra. *Educ. Admin. & Supervis.*, 1926, **12**, 512–518.

11. BASS, B. M. An analysis of the leaderless group discussion. *J. appl. Psychol.*, 1949, **33**, 527–633.

12. BASS, B. M., & NORTON, F. M. Group size and leaderless discussions. *J. appl. Psychol.*, 1951, **35**, 397–400.

13. BASS, B. M., & WHITE, D. L., JR. Validity of leaderless group discussion observers' descriptive and evaluative ratings for the assessment of personality and leadership status. *Amer. Psychologist*, 1950, **5**, 311–312. (Abstract)

14. BAVELAS, A. Unpublished manuscript. Cited in K. Lewin, Frontiers in group dynamics. *Hum. Relat.*, 1947, **1**, 5–41.

15. BAVELAS, A. A mathematical model for group structures. *Applied Anthropol.*, 1948, **7**, 16–30.

16. BAVELAS, A. Communication patterns in task-oriented groups. In D. Lerner & H. Lasswell (Eds.), *The policy sciences; recent developments in scope and method.* Stanford: Stanford Univer. Press, 1951. Pp. 193–202.

17. BAVELAS, A., & BARRETT, D. An experimental approach to organizational communication. *Personnel*, 1951, **27**, 367–371.

18. BECHTEREW, W., & DE LANGE, M. Die Ergebnisse des Experiments auf dem Gebiete der kollestiven Reflexologie. *Zsch. f. angew. Psychol.*, 1924, **24**, 305–344.

19. BELL, G. B., & FRENCH, R. I. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 764–767.

20. BERRIEN, F. K. Attempts to measure attitudinal changes as a consequence of permissive discussions. *Amer. Psychologist*, 1950, **5**, 246–247. (Abstract)

21. BETTELHEIM, B., & SYLVESTER, EMMY. Therapeutic influence of the group on the individual. *Amer. J. Orthopsychiat.*, 1947, **27**, 684–692.

22. BION, W. R., & RICKMAN, J. Intragroup tensions in therapy; their study as the task of the group. *Lancet*, 1943, **245**, 678–681.

23. BONNEY, MERL E. A sociometric study of the relationships of some factors to mutual friendships on the elementary, secondary and college levels. *Sociometry*, 1946, **9**, 21–47.

24. BOVARD, E. W., JR. Social norms and the individual. *J. abnorm. soc. Psychol.*, 1948, **43**, 62–69.

25. BOVARD, E. W., JR. Group structure and perception. *J. abnorm. soc. Psychol.*, 1951, **46**, 398–405.

26. BOVARD, E. W., JR. The experimental production of interpersonal affect. *J. abnorm. soc. Psychol.*, 1951, **46**, 521–528.

27. BROWN, IDA S. Training university students in group development and experiment. Unpublished doctor's dissertation, Univer. of California, 1950.

28. BURTT, H. E. Sex differences in the effect of discussion. *J. exp. Psychol.*, 1920, **3**, 390–395.

29. CARTER, L. F., HAYTHORN, W., & HOWELL, MARGARET. A further investigation of the criteria of leadership. *J. abnorm. soc. Psychol.*, 1950, **45**, 350–358.

30. CARTER, L. F., HAYTHORN, W., SHRIVER, BEATRICE, & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589–595.

31. CARTER, L. F., & NIXON, MARY. An investigation of the relationship between four criteria of leadership ability for three different tasks. *J. Psychol.*, 1949, **27**, 245–261.

32. CARTER, L. F., & NIXON, MARY. Ability, perceptual, personality and interest factors associated with different criteria of leadership. *J. Psychol.*, 1949, **27**, 277–388.

33. CATTELL, R. B. Determining syntality dimension as a basis for morale and leadership measurement. In H. Guetzkow (Ed.), *Groups, leadership and men: research in human relations.*

Pittsburgh: Carnegie Press, 1951. Pp. 16–27.

34. CHAPPLE, E. D. The interaction chronograph: its evolution and present application. *Personnel*, 1949, 25, 295–307.

35. COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, 1, 512–532.

36. COFFEY, H., FREEDMAN, M., LEARY, T., & OSSORIO, A. (Eds.) Community service and social research: group psychotherapy in a church program. *J. soc. Issues*, 1950, 6, No. 1.

37. DASHIELL, J. F. Experimental studies of the influence of social situations on the behavior of individual human adults. In C. Murchison (Ed.), *Handbook of social psychology.* Worcester, Mass.: Clark Univer. Press, 1935. Pp. 1097–1158.

38. DAVIS, A. The motivation of the underprivileged worker. In W. F. Whyte (Ed.), *Industry and society.* New York: McGraw-Hill, 1946. Pp. 84–106.

39. DEUTSCH, M. An experimental study of the effects of cooperation and competition upon group process. *Hum. Relat.*, 1949, 2, 199–231.

40. DEUTSCH, M. Task structure and group process. *Amer. Psychologist*, 1951, 6, 324–325. (Abstract)

41. ELKES, REGINA. Group casework experiment with mothers of children with cerebral palsy. *J. soc. Casework*, 1947, 28, 95–101.

42. EVANS, J. T. Some measurements of interaction in two therapy groups. Unpublished doctor's dissertation, Harvard Univer., 1950.

43. FAW, V. A. A psychotherapeutic method of teaching psychology. *Amer. Psychologist*, 1949, 4, 104. (Abstract)

44. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups: a study of human factors in housing.* New York: Harper, 1950.

45. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, 46, 92–99.

46. FLANDERS, N. A. Personal-social anxiety as a factor in experimental learning situations. *J. educ. Res.*, 1951, 45, 100–110.

47. FOURIEZOS, N. T., HUTT, M. L., & GUETZKOW, H. Measurement of self-oriented needs in discussion groups. *J. abnorm. soc. Psychol.*, 1950, 45, 682–690.

48. FOX, J. B., & SCOTT, J. F. *Absenteeism: management's problem.* Boston: Graduate School of Business Administration, Harvard Univer., 1943.

49. FRENCH, J. R. P., JR. The disruption and cohesion of groups. *J. abnorm. soc. Psychol.*, 1941, 36, 361–377.

50. FRENCH, J. R. P., JR. Organized and unorganized groups under fear and frustration. *Univer. Ia. Stud. Child Welf.*, 1944, 20, No. 409, 229–308.

51. FRENCH, J. R. P., JR., KORNHAUSER, A., & MARROW, A. Conflict and cooperation in industry. *J. soc. Issues*, 1946, 2, 2–55.

52. FRENCH, J. R. P., JR., & ZANDER, A. The group dynamics approach. In A. Kornhauser (Ed.), *Psychology of labor-management relations.* Champaign, Ill.: Industrial Relations Research Assoc., 1949. Pp. 71–80.

53. FRENCH, R. L. Verbal output and leadership status in initially leaderless discussion groups. *Amer. Psychologist*, 1950, 5, 310–311. (Abstract)

54. FRENCH, R. L. Sociometric status and individual adjustment among naval recruits. *J. abnorm. soc. Psychol.*, 1951, 46, 64–71.

55. FULTON, RUTH E. Relationship between teammate status and measures of skill in volleyball. *Res. Quart. Amer. Ass. Hlth*, 1950, 21, 274–275.

56. GIBB, C. A. The emergence of leadership in small temporary groups of men. Unpublished doctor's dissertation, Univer. of Illinois, 1949.

57. GIBB, J. R., PLATTS, GRACE N., & MILLER, LORRAINE F. Dynamics of participative groups. Lithographed manuscript. Univer. of Colorado, 1951.

58. GOODENOUGH, FLORENCE L. Interrelationships in the behavior of young children. *Child Develpm.*, 1930, 1, 29–48.

59. GORDON, KATE. A study of esthetic judgments. *J. exp. Psychol.*, 1923, 6, 36–43.

60. GORDON, KATE. Group judgments in the field of lifted weights. *J. exp. Psychol.*, 1924, 7, 398–400.

61. GREEN, N. E. Verbal intelligence and effectiveness of participation in group discussion. *J. educ. Psychol.*, 1950, 41, 440–445.

62. GROSS, L. An experimental study of the validity of the non-directive method of teaching. *J. Psychol.*, 1948, 26, 243–248.

63. GROSSER, D., POLANSKY, N., & LIP-

PITT, R. A laboratory study of behavioral contagion. *Hum. Relat.*, 1951, 4, 115–142.

64. GURNEE, H. Group interaction in a learning situation. *Amer. Psychologist*, 1948, 3, 270. (Abstract)

65. GUTHE, C. E. (Chm.) Manual for the study of food habits; report of the Committee on Food Habits. *Bull. nat. Res. Coun., Wash.*, 1945, No. 111.

66. GYR, J. Analysis of committee member behavior in four cultures. *Hum. Relat.*, 1951, 4, 193–202.

67. HEISE, G. A., & MILLER, G. A. Problem solving by small groups using various communication nets. *J. abnorm. soc. Psychol.*, 1951, 46, 327–336.

68. HEMPHILL, J. K. Situational factors in leadership. *Ohio St. Univer. Stud. Bur. Educ. Res. Monogr.*, 1949, No. 32.

69. HEMPHILL, J. K. Relations between the size of the group and the behavior of "superior" leaders. *J. soc. Psychol.*, 1950, 32, 11–22.

70. HENRY, W. E., & GUETZKOW, H. Group projection sketches for the study of small groups. *J. soc. Psychol.*, 1951, 33, 77–102.

71. HEYNS, R. W. Functional analysis of group problem-solving behavior. Unpublished manuscript, Conference Research Project, Univer. of Michigan, 1948.

72. HEYNS, R. W. Effects of variation in leadership on participant behavior in discussion groups. Unpublished doctor's dissertation, Univer. of Michigan, 1949.

73. HOLLINGSHEAD, A. B. Cultural factors in the selection of marriage mates. *Amer. sociol. Rev.*, 1950, 15, 619–627.

74. HOLLINGWORTH, LETA S. *Gifted children: their nature and nurture.* New York: Macmillan, 1926.

75. HOMANS, G. C. *The human group.* New York: Harcourt Brace, 1950.

76. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, 4, 57–76.

77. HORSFALL, A. B., & ARENSBERG, C. M. Teamwork and productivity in a shoe factory. *Hum. Organization*, 1949, 8, 13–25.

78. HUNT, J. McV., & SOLOMON, R. L. The stability and some correlates of group status in a summer camp of young boys. *Amer. J. Psychol.*, 1942, 55, 33–45.

79. HUSBAND, R. A statistical comparison of the efficacy of large lecture versus smaller recitation sections upon the achievement in general psychology. *Amer. Psychologist*, 1949, 4, 216. (Abstract)

80. JAQUES, E. Interpretive group discussion as a method of facilitating social change. *Hum. Relat.*, 1948, 1, 533–549.

81. JENKINS, D. H. Feedback and group self-evaluation. *J. soc. Issues*, 1948, 4, 50–60.

82. JENKINS, W. O. A review of leadership studies with particular reference to military problems. *Psychol. Bull.*, 1947, 44, 54–79.

83. JENNESS, A. Social influences in the change of opinion; the role of discussion in changing opinion regarding a matter of fact. *J. abnorm. soc. Psychol.*, 1932, 27, 29–34, 279–296.

84. JENNINGS, HELEN H. *Leadership and isolation: a study of personality in interpersonal relations.* (2nd Ed.) New York: Longmans, Green, 1950.

85. JOEL, W., & SHAPIRO, D. A genotypical approach to the analysis of personal interaction. *J. Psychol.*, 1949, 28, 9–17.

86. JOLLES, I. An experiment in group guidance. *J. soc. Psychol.*, 1946, 23, 55–60.

87. JONES, H. E. Experimental studies of college teaching. *Arch. Psychol.*, 1923, 10, No. 68.

88. KELLEY, H. H. Communication in experimentally created hierarchies. *Hum. Relat.*, 1951, 4, 39–56.

89. KLOPFER, W. G. The efficacy of group therapy as indicated by group Rorschach records. *Rorschach Res. Exch.*, 1945, 9, 207–209.

90. KLUGMAN, S. F. Cooperative versus individual efficiency in problem-solving. *J. educ. Psychol.*, 1944, 35, 91–100.

91. KLUGMAN, S. F. Group and individual judgments for anticipated events. *J. soc. Psychol.*, 1947, 26, 21–28.

92. LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, 46, 38–50.

93. LEAVITT, H. J., & MUELLER, R. A. H. Some effects of feedback on communication. *Hum. Relat.*, 1951, 4, 401–410.

94. LEMANN, T. B. An empirical investigation of group characteristics as revealed in sociometric patterns and

personality ratings. Unpublished bachelor's dissertation, Harvard University., 1949.

95. LEWIN, K. Forces behind food habits and methods of change. In C. E. Guthe (Chm.), The problem of changing food habits; report of the Committee on Food Habits, 1941–1943. *Bull. nat. Res. Counc.*, Wash., 1943, No. 108, 35–65.

96. LEWIN, K., LIPPITT, R., & WHITE, R. K. Patterns of aggressive behavior in experimentally created "social climates." *J. soc. Psychol.*, 1939, 10, 271–299.

97. LINDZEY, G., & RIECKEN, H. W. Inducing frustration in adult subjects. *J. consult. Psychol.*, 1951, 15, 18–23.

98. LIPPITT, R. Training in community relations; a research exploration toward new group skills. New York: Harper, 1949.

99. LIPPITT, R., & RADKE, M. New trends in the investigation of prejudice. *Ann. Amer. Acad. polit. soc. Sci.*, 1946, 244, 167–176.

100. LIPPITT, R., & WHITE, R. K. The "social climate" of children's groups. In R. Barker, J. Kounin, & H. Wright (Eds.), *Child development and behavior.* New York: McGraw-Hill, 1943. Pp. 485–508.

101. LIPPITT, R., & WHITE, R. K. An experimental study of leadership and group life. In T. Newcomb & E. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1947. Pp. 315–330.

102. LORENZ, E. Zur psychologie der industriellen Gruppenarbeit. *Zsch. f. angew. Psychol.*, 1933, 45, 1–45.

103. MAAS, H. S. Personal and group factors in leaders' social perception. *J. abnorm. soc. Psychol.*, 1950, 45, 54–63.

104. MAIER, N. R. F. The quality of group decisions as influenced by the discussion leader. *Hum. Relat.*, 1950, 3, 155–174.

105. MALLER, J. B. Cooperation and competition; an experimental study in motivation. *Teach. Coll. Contrib. Educ.*, 1929, No. 384.

106. MALONE, T. P. Analysis of the dynamics of group psychotherapy based on observation in a twelve-month experimental program. *J. Pers.*, 1948, 16, 245–277.

107. MARRIOTT, R. Size of working group and output. *Occup. Psychol.*, Lond., 1949, 23, 47–57.

108. MARROW, A. J. Group dynamics in industry; implications for guidance and personnel workers. *Occupations*, 1948, 26, 472–476.

109. MARROW, A. J., & FRENCH, J. R. P., JR. Changing a stereotype in industry. *J. soc. Issues*, 1945, 1, 33–37.

110. MARSTON, W. M. Studies in testimony. *J. crim. Law Criminol.*, 1924, 15, 5–31.

111. MAYO, E. *The human problems of an industrial civilization.* New York: Macmillan, 1933.

112. MAYO, E., & LOMBARD, G. F. F. *Teamwork and labor turnover in the aircraft industry of Southern California.* Boston: Graduate School of Business Administration, Harvard Univer., 1944.

113. McCANDLESS, B. R. Changing relationships between dominance and social acceptability during group democratization. *Amer. J. Orthopsychiat.*, 1942, 12, 529–536.

114. MEREI, F. Group leadership and institutionalization. *Hum. Relat.*, 1949, 2, 23–39.

115. MERRILL, BARBARA. A measurement of mother-child interaction. *J. abnorm. soc. Psychol.*, 1946, 41, 37–49.

116. MINTZ, A. Non-adaptive group behavior. *J. abnorm. soc. Psychol.*, 1951, 46, 150–159.

117. MISHLER, E. Ascendant and submissive members and leaders: their interaction in group discussion. Unpublished manuscript, Conference Research Project, Univer. of Michigan, 1950.

118. MOEDE, W. Die Richtlinien der Leistungs-Psychologie. *Indus. Psychotech.*, 1927, 4, 193–209.

119. MORENO, J. L. *Who shall survive? A new approach to the problem of human interrelations.* New York: Beacon House, 1934.

120. MÜNSTERBERG, H. *Psychology and social sanity.* Garden City, N. Y.: Doubleday, 1914.

121. NORFLEET, BOBBIE. Interpersonal relations and group productivity. *J. soc. Issues*, 1948, 4, 66–69.

122. NORTHWAY, MARY BLOSSOM, & WIGDOR, B. T. Rorschach patterns related to the sociometric status of school children. *Sociometry*, 1947, 10, 186–199.

123. OFFICE OF STRATEGIC SERVICES ASSESSMENT STAFF. *Assessment of men.* New York: Rinehart, 1948.

124. OSSORIO, A. G., & LEARY, T. Patterns of social interaction and their relation to personality structure. *Amer. Psychologist*, 1950, 5, 303. (Abstract)

125. PEPINSKY, H. B. Brief group psychotherapy and role and status: a case study. *Amer. Psychologist*, 1949, 4, 294. (Abstract)

126. PERKINS, H. V. Climate influences group learning. *J. educ. Res.*, 1951, 45, 115–119.

127. PLANK, R. An analysis of a group therapy experiment. *Hum. Organization*, 1951, 10 (3), 5–21.

128. POLANSKY, N., LIPPITT, R., & REDL, F. The use of near-sociometric data in research on group treatment processes. *Sociometry*, 1950, 13, 39–62.

129. POLANSKY, N., LIPPITT, R., & REDL, F. An investigation of behavioral contagion in groups. *Hum. Relat.*, 1950, 3, 319–348.

130. PRESTON, M. G., & HEINTZ, R. K. Effects of participatory versus supervisory leadership on group judgment. *J. abnorm. soc. Psychol.*, 1949, 44, 345–355.

131. RADKE, M., & KLISURICH, D. Experiments in changing food habits. *J. Amer. dietetics Ass.*, 1947, 23, 403–409.

132. RADU, I. Cooperaria in activitatea intellectuala. *Rev. Psihol.*, 1948, 11, 107–116.

133. REHAGE, K. J. A comparison of pupil-teacher planning and teacher directed procedures in eighth grade social studies classes. *J. educ. Res.*, 1951, 45, 111–115.

134. RIDDLE, E. H. Aggressive behavior in a small social group. *Arch. Psychol.*, 1925, No. 78.

135. ROETHLISBERGER, F. J., & DICKSON, W. J. *Management and the worker: social versus technical organization in industry.* Cambridge, Mass.: Harvard Univer. Press, 1939.

136. ROSEBOROUGH, MARY E. The effect of group composition upon students' interaction and academic achievement. Unpublished master's dissertation, Univer. of Toronto, 1949.

137. ROSENTHAL, D., & COFER, C. N. The effect on group performance of an indifferent and neglectful attitude shown by one group member. *J. exp. Psychol.*, 1948, 38, 568–577.

138. SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, 46, 190–207.

139. SCHACHTER, S., ELLERTSON, N., McBRIDE, DOROTHY, & GREGORY, DORIS. An experimental study of cohesiveness and productivity. *Hum. Relat.*, 1951, 4, 229–238.

140. SCHONBAR, ROSALEA ANN. The modification of judgments in a group situation. *J. exp. Psychol.*, 1947, 37, 69–80.

141. SHAW, MARJORIE E. A comparison of individuals and small groups in the rational solution of complex problems. *Amer. J. Psychol.*, 1932, 44, 491–504.

142. SHERIF, M. *The psychology of social norms.* New York: Harper, 1936.

143. SHILS, E. *The present state of American sociology.* Glencoe, Ill.: Free Press, 1948.

144. SIMMEL, G. The number of members as determining the sociological form of the group. *Amer. J. Sociol.*, 1902, 8, 1–46, 158–196.

145. SMITH, M. Some factors in friendship selections of high school students. *Sociometry*, 1944, 7, 303–310.

146. SMITH, S. L. Communication pattern and the adaptability of task-oriented groups: an experimental study. Cited in A. Bavelas, Communication patterns in task-oriented groups. In D. Lerner & H. Lasswell (Eds.), *The policy sciences; recent developments in scope and method.* Stanford, Calif.: Stanford Univer. Press, 1951. Pp. 193–202.

147. SOUTH, E. B. Some psychological aspects of committee work. *J. appl. Psychol.*, 1927, 11, 348–368, 437–464.

148. SPENCE, R. B. Lecture and class discussion in teaching educational psychology. *J. educ. Psychol.*, 1928, 19, 454–462.

149. STATON, T. F. An analysis of the effect of individuals on seminar discussion. *Amer. Psychologist*, 1948, 3, 267. (Abstract)

150. STEINZOR, B. The spatial factor in face-to-face discussion groups. *J. abnorm. & soc. Psychol.*, 1950, 45, 552–555.

151. STENDLER, CELIA, DAMRIN, DORA, & HAINES, ALEYNE C. Studies in cooperation and competition: I. The effect of working for group and individual rewards on the social climate of children's groups. *J. genet. Psychol.*, 1951, 79, 173–197.

152. STERLING, T. D., & ROSENTHAL, B. G. The relationship of changing leadership and followership in a group to the changing phases of group activity. *Amer. Psychologist*, 1950, 5, 311. (Abstract)

153. STOGDILL, R. M. Personal factors associ-

ated with leadership: a survey of the literature. *J. Psychol.*, 1948, **25**, 35–71.

4. STRODTBECK, F. L. Husband-wife interaction over revealed differences. *Amer. sociol. Rev.*, 1951, **16**, 468–473.

55. STROOP, J. R. Is the judgment of the group better than that of the average member of the group? *J. exp. Psychol.*, 1932, **15**, 550–562.

56. SWANSON, G. E. Some effects of member object-relationships on small groups. *Hum. Relat.*, 1951, **4**, 355–380.

57. SWANSON, G. E. Some problems of laboratory experiments with small populations. *Amer. sociol. Rev.*, 1951, **16**, 349–357.

158. TAGIURI, R. Relational analysis, an extension of sociometric method. Unpublished doctor's dissertation, Harvard Univer., 1951.

159. THELEN, H. A., & WHITHALL, J. Three frames of reference: the description of climate. *Hum. Relat.*, 1949, **2**, 159–176.

160. THIBAUT, J. An experimental study of the cohesiveness of underprivileged groups. *Hum. Relat.*, 1950, **3**, 251–278.

161. THOMPSON, W. R., & NISHIMURA, RHODA. Some determinants of friendship. *Amer. Psychologist*, 1950, **5**, 309. (Abstract)

162. VAN ZELST, R. H. Worker popularity and job satisfaction. *Personnel Psychol.*, 1951, **4**, 405–412.

163. WATSON, G. B. Do groups think more efficiently than individuals? *J. abnorm. soc. Psychol.*, 1928, **23**, 328–336.

164. WHITEHEAD, T. N. *Leadership in a free society: a study in human relations based on an analysis of present day industrial civilization.* Cambridge, Mass.: Harvard Univer. Press, 1937.

165. WHYTE, W. F. *Street corner society: the social structure of an Italian slum.* Chicago: Univer. of Chicago Press, 1943.

166. WILLERMAN, B. Group decision and request as a means of changing food habits. Unpublished manuscript, Committee of Food Habits, National Research Council, Washington, 1943.

167. WISPÉ, L. G. Evaluating section teaching methods in the introductory course. *J. educ. Res.*, 1951, **45**, 161–186.

168. WRIGHT, M. E. The influence of frustration upon the social relations of young children. *Charact. & Pers.*, 1943, **12**, 111–122.

169. ZELENY, L. D. Selection of compatible flying partners. *Amer. J. Sociol.*, 1947, **52**, 424–431.

ALEXANDER, FRANZ, & ROSS, HELEN. (Eds.) *Dynamic psychiatry*. Chicago: Univer. of Chicago Press, 1952. Pp. xii+578. $10.00.

The announcement of this book is likely to spur some scientific minds, prudent though they may be, to a gallop of expectations. It is introduced as the promised fulfillment of a current need: "the most complete picture possible at this time of the fundamental ideas of dynamic psychiatry." The application of psychoanalytic theories to the understanding of neuropsychiatric disorders has never been fully and adequately represented in one concise volume. The title of this book—in which the word "dynamic" has been substituted for "psychoanalytic"—speaks for an intention to do just this and more: to assess the utility of all promising directional concepts and propositions regardless of their ancestry. The title suggests that the editors realize that Freud was not the parent or grandparent of *every* relevant dynamical principle, and that their compilation, consequently, will extend beyond the prescribed limits of a parochial textbook. The list of authors is a dazzle of eminence: some of America's foremost psychiatrists—seven from the Chicago Institute for Psychoanalysis —and, in their company, a distinguished anthropologist and psychologist, Margaret Mead and David Shakow. That Franz Alexander, one of the most imaginative and wide-ranging of analysts, was the guiding light of this enterprise is close to a guarantee of large and unprejudiced dimensions. Finally, there is an inviting table of contents and an agreeable format. Here, surely, is a summation of signals that is more than

enough to kick off the salivary reflex and whet one's appetite for the fare between these covers.

The reader's satisfaction with *Dynamic Psychiatry* can be more confidently predicted, however, if he will shut his senses to such signals and, instead, perform a little ritual of negative magic. First of all, he should prepare his cognitive machinery for a book which, like many another, begins and ends with Freud.

As Brosin says in the last chapter: "The system of ideas worked out and presented to the world by one man, Sigmund Freud (1856–1939), has exercised so profound and far-reaching an influence on the minds of men that it may well be ranked among the great new ideas which, like Darwin's theory of evolution, have shaped the course of history." With this judgment and prophecy most of us psychiatrists and psychologists, I presume, would unhesitatingly agree, especially since by this agreement we provide ourselves with defensible ground for a high evaluation of our profession. None could be more deserving of our everlasting gratitude than the man whose penetrating observations and formative reflections have made the once inconsequential disciplines of descriptive psychiatry and academic psychology consequential, and who, in so doing, has provided us with new and dynamic modes of thought for the analysis and clarification of elemental human problems. What could be more timely, if the contemporary world is approximately what it seems to be—*in extremis*, threatened and disoriented, a great tangle of elemental conflicts crying for illumination and decision? Without Freud, today's psychiatrists

304

and psychologists would have very little, if anything, to say that is applicable in this crisis; but *with* Freud, there is a slight possibility, despite his overshadowing pessimism, that they may contribute to the construction and verification of a healing ideology which, through child training and education, might shape, as Brosin suggests, the course of history.

How Freud's thoughts have influenced or might influence the course of history does not become apparent here, but what *does* become apparent is the exclusive degree to which they have shaped the course of this book, *Dynamic Psychiatry*. According to my count—for those who respect numbers—within the first two pages of each of the sixteen chapters (32 pages in all) Freud's name is mentioned more than fifty times and no other name from Aristotle down is mentioned more than three or four times. There are arrays of facts which show how much social scientists have learned from Freud, but few evidences that Freudians have learned anything or have anything to learn (which is it?) from the other social sciences. This might be taken as one of countless proofs of Freud's towering genius which "doth bestride the narrow world" of psychological theory "like a Colossus," so that "we petty men" can do nothing "under his huge legs," but "peep about," hoping against hope to find an unnoticed little fact or two; and/or this might be taken as one of countless proofs of Freudiolatry, or an Omniscient Father complex, which, if too compelling, would be the nemesis of creativity, since, in Whitehead's words, "a science which hesitates to forget its founders is lost." Freudiolatry, as an imprisoner of thought, is nowhere more painfully conspicuous than in the area of motivation (in-

stinct, need, drive), where the analysts' loyalty to their charismatic author confines them to the dramatic, but wholly inadequate, dichotomy of Empedocles (love and hate, libido and aggression, Eros and Thanatos). It is necessary to heap rationalization on rationalization in order to subsume under these two headings the galaxy of expressive and purposive dispositions which constitute human nature—from anoxemia, thirst, hunger, and excretion, the sucking reflex and the maternal drive, the dread of injury, ignominy, and death, the craving for property, power, and prestige, to the mental needs for the creation and expression of comic, aesthetic, scientific, and moral-legal forms. The book is mute in respect to the here-pertinent conceptions of McDougall, Tolman, Adler, Cattell, and others.

Secondly, the prospective reader should prepare his mind for a compilation of more or less independent essays on a variety of topics, rather than an integrated survey of "the fundamental ideas of dynamic psychiatry." Psychoanalytic concepts are expounded in different ways in different places, but no statements of them compare in clarity and distinctness to those of Hartmann, Kris, and Loewenstein. Furthermore, in one crucial area, neither the comprehensiveness nor the organization of this book approaches the level of Fenichel's *Outline of Clinical Psychoanalysis*. Only twenty-three pages (about 4 per cent of the total) are devoted to a systematic account of "neuroses, behavior disorders, and perversions."

Having steeled himself against disappointments of these sorts the reader will be in a position to enjoy and to be greatly instructed by a number of excellent articles. Two

chapters—one by Therese Benedek and the other by Margaret W. Gerard—taken together, provide an admirable picture of normal and abnormal developments in childhood and adolescence as seen through Freudian and Alexandrian spectacles. I, for one, was rewarded by reading the carefully considered sentences of John C. Whitehorn. In the "Principles of Psychiatric Treatment" Maurice Levine presents a clear and concise outline of the tried practices of the therapist and the tendencies in the patient which commonly promote and those which commonly impede the curative process. Alexander is at his best with Thomas S. Szasz in a brief 26-page review of the psychosomatic approach in medicine. In "Some Relationships between Social Anthropology and Psychiatry" Margaret Mead is as interesting as ever; and David M. Levy is perfect, according to my scales, in his informative, judicious, and well-written "Animal Psychology in its Relation to Psychiatry." Altogether there is much substantial nourishment to be extracted from these pages.

HENRY A. MURRAY
Harvard University

DEESE, JAMES. *The psychology of learning.* New York: McGraw-Hill, 1952. Pp. ix+398. $5.50.

In the preface to this book, Deese tells us that he has tried to bring together "all the divergent interests in learning" within a text that is "by and large empirical in approach," of "strong behavioristic flavor," and intended primarily for college juniors and seniors. This is a fair statement of what he has done. Except for *motor skills*, no major interest is absent; the account is satisfactorily empirical, the flavor objective, and the writing geared to undergraduates in all but a few paragraphs (e.g., in discussing "Current Problems . . .," pp. 343–353).

In the Introduction, however, he promises a book with a theoretical framework of such "wide application" that "a great many examples of learning will be covered with a relatively economical set of principles and concepts" (p. 3). He leads us to expect that the first third of the book (Chs. 2–7, Reinforcement and Learning, The Nature of Extinction, Stimuli in Learning and the Process of Discrimination, Motivation and Learning, Negative Reinforcement and Punishment, Serial Learning and the Chaining of Responses) will be truly basic to chapters in the second third (Chs. 9–12, Retention, Forgetting, Transfer of Training, and Efficiency in Human Learning) and not unrelated to such later chapters as Thinking and Problem Solving (Ch. 13) or Emotion and Conflict (Ch. 15). In this he lets us down. We get, instead, an unrepresentative and badly edited account of reinforcement theory; a cluster of McGeoch-type chapters on verbal learning one of which (Ch. 8, Factors Affecting Rate of Complex Learning) should have been spread among three others (Chs. 7, 9, and 12), and two of which (Retention and Forgetting) are kept apart on the assumption that forgetting is an intentionally or experimentally induced loss of retention; and a final section of five chapters, four of which (Chs. 13, Thinking . . .; 14, Learning and the Nature of the Learner; 15, Emotion and Conflict; 16, Physiological Problems in Learning) are little more than 1952 models of their 1940 Hilgard-and-Marquis counterparts. No new systematic relations appear in all this; and some old familiar ones are either absent or seriously blurred.

Generalization, for example, is introduced in Chapter 4. It properly carries the burden of transfer in Chapter 11. In Chapter 7, it accounts for the anticipatory and perseverative errors that produce the serial position effect in Hull's linear maze, and the suggestion is made (p. 141) that it will look after remote associations in rote learning. But, in Chapter 8, the serial position effect in both linear-maze nonsense-syllable learning is attributed to remote associations, as well as to the Pavlov-Lepley-Hull inhibitory process (p. 164). Later, in the same chapter, the position effect is cited as an example of intratask interference, which is in turn said to arise from intratask "similarity" (p. 168). Finally, generalization is mentioned but once (p. 198) in discussing retroaction (where "similarity" is again used), and not at all in treating "insightful problem solving" ("a special case of . . . transfer") or concept formation.

The first six chapters will give Deese's readers the most trouble. Here the influence of Skinner and Hull predominates, but the viewpoint offered is that of neither. Reinforcement theorists will be disturbed to find that for extinction to occur an organism must "*perceive* the absence of reinforcement" (p. 62); that "the operational definition of drives . . . will not do for human learning" (p. 98); that negative reinforcement is "painful stimulation for *not* doing something," and is applied to "poorly motivated" behavior, whereas punishment is "painful stimulation for *doing* something" and is applied to "highly motivated" behavior (p. 111) and that "generalization [in serial learning] could be simply on the basis of the time sense: that is, errors could occur because the subject 'loses his place' " (p. 141). Later

on in the book they may also be distressed by the use of such explanatory decoys as "ego involvement" (pp. 178, 180), "set" (pp. 181, 193), and "*Einstellung*" (p. 262). Other readers, even undergraduates, will be irritated by numerous slips that should have been picked up in the editor's office. For example, there is the assertion that the "percentage [*sic*] of conditioned responses becomes more and more frequent as we increase the number of reinforcements" (p. 16); there is the use of "secondary reinforcement" in the last two sentences of page 110, when "negative reinforcement" must have been meant; and there are several cases in which the textual reference to figures (e.g., Figs. 2, 15, 44, and 49) is inaccurate in some detail, trivial or otherwise.

Nevertheless, Deese's book has points of real merit. The narrative marches at a good clip, reader interest is generally maintained (at least *this* reader's interest was), and there is a welcome freshness in the selection of researches reported. Several chapters are of first-rate quality (e.g., the ones on chaining and transfer) and others don't miss the mark by far. For that matter, the whole book points in the right direction and it might have come off successfully if another year had passed before it went to press.

FRED S. KELLER

Columbia University

DAVIDSON, HENRY A. *Forensic psychiatry.* New York: Ronald Press, 1952. Pp. viii+398. $8.00.

Many a lawyer seems to confuse the fields of psychology, psychiatry, and psychoanalysis. In general, he is likely to believe that all psychologists deal with problems of abnormality, or else that they spend their spare

moments chasing white rats through mazes.

Equally confusing is the picture which many a psychologist or psychiatrist has of the field of law. Most psychologists and psychiatrists, along with other nonlawyers, tend to think of all law as being *criminal* law. They also fall into the error of thinking of practically all law practice as involving trial work and courtroom problems.

The distinguished author of the present volume commits neither of these errors. He is well aware of the civil as well as the criminal problems in the field of law. He is conversant with problems of legal psychology which do not center on trial procedures.

Dr. Davidson is Chief of the Program Analysis and Planning Section of the Psychiatry and Neurology Division of the Veterans Administration. His book on forensic psychiatry demonstrates that he has had a wealth of experience in various fields of psychiatry. The present volume gives the reader pretty good insight into courtroom situations so far as the expert witness, especially the psychiatrist, is concerned. At the same time it deals with other important areas as well—problems of competency, malingering, the last will and testament, appraisal of the sex offender, evaluation of personal injury, etc. Essentially this is a manual for physicians and psychiatrists, to serve them as a psychiatric legal guide. Thus it enables specialists in medical and clinical practice to gain a better understanding of the psychiatric aspects of the law. Certainly other professional workers, especially clinical psychologists and social workers, will also find certain chapters of particular use. The chapter on malingering would be of interest to any clinician.

The first half of the book deals with the contents of forensic psychiatry. The latter half describes the tactics of testimony, and this includes some very practical advice to the medical witness as to what to expect and how to be prepared for various courtroom situations. Of particular interest to anyone with an interest in semantics is Chapter 21, The Translation of Technical Terms.

The Appendix is unique. Among other things it contains a legal lexicon for doctors, so that they will have some notion of what is meant by such terms as contributory negligence, exemplary damage, hearsay testimony, and self-serving statements. The psychiatric glossary for lawyers, in turn, gives the man of law some impression of what the psychiatrist has in mind when he uses such words as ambivalence, compulsion neurosis, etiology, psychopathic, and the like. It is unfortunate that a psychologist is defined as "a nonphysician trained in the understanding of mental mechanisms"; and it is equally unfortunate that no definition is included of a psychiatrist.

A significant contribution of the book is a detailed proposal for a model act of the certification of the mentally ill. This should be useful in many different jurisdictions.

It is difficult to find fault with this work. It is easier and more accurate to say that one can learn a good deal from studying it.

STEUART HENDERSON BRITT
*Needham, Louis and Brorby, Inc.,
Chicago, Ill.*

RIESEN, AUSTIN H., & KINDER, ELAINE. *The postural development of infant chimpanzees. A comparative and normative study based on the Gesell behavior examination.* New Haven: Yale Univer. Press. 1952. Pp. xx+204. $5.00.

This volume is one of a projected series to be published in connection with the Infant Studies Program which has been in progress at the Yerkes Laboratories of Primate Biology since 1939. Chimpanzee infants have been isolated from their mothers and reared in an experimental nursery under carefully controlled conditions while various aspects of their development have been charted. Riesen and Kinder report on postural development in 14 apes during the first year of life.

The tests were selected from the Gesell and Thompson examination. They included the postural situations (supine, prone, pulled to sitting, sitting, and standing) and some data on the grasping response obtained in the Rod situation. The evidence is presented in considerable detail (18 tables and 117 graphs). At each step, the performance of the chimpanzee is compared with that of human children of comparable age as determined by Gesell and Thompson.

The development of a postural repertoire in human and chimpanzee infants is quite similar. In both species the individual gradually gains control of the head, loses some neonatal reflexes, "learns" to roll over, to sit, and to stand and walk upright. Independent movements of arms, legs, hands, and feet change from month to month. Developmental similarities between chimpanzees and humans are closer for head and general body postures than for postures of the arm, leg, or hand. The ages of appearance of the developmental items average one-third earlier for chimpanzee infants in those cases where a valid basis for comparison is available. On a few items the human infant shows an earlier age of development than the ape. The investigators devised a postural schedule for chimpanzees, con-

sisting of 60 selected items which showed a split-half reliability of .95. When this schedule was applied to infants left with their mothers or reared in a human home these animals were found to fall within the range of the nursery group. In conclusion the authors critically examine several theories and concepts of behavioral development and find that some of these are not supported by their data. Other concepts, such as that of the assumption of cortical control, that of alternation of extensor and flexor dominance, and the theory of of correspondence, are in agreement with Riesen and Kinder's findings.

This extremely detailed and at times highly technical analysis of postural development is a scholarly achievement and an important addition to the literature on genetic psychology. With the exception of Marion Hine's monograph on the macaque there is no comparable work of equal calibre. It is to be hoped that subsequent reports growing out of the Infant Studies Program will meet the high standards set by this opening publication in the series.

FRANK A. BEACH

Yale University

GORLOW, L., HOCH, E. L., & TELSCHOW, E. F. *The nature of nondirective group psychotherapy.* New York: Teachers College, Columbia Univer., 1952. Pp. xii+143. $3.25.

This book is a report of statistical analyses of data secured in nondirective group psychotherapeutic sessions conducted at Columbia University. The three authors served as therapists in the study. Its primary value is as a pioneering research project which demonstrates the applicability of research analyses to some aspects of group psychotherapy. However, the conclusions reached were based on data secured from only seventeen

clients, all graduate students, and only three therapists. Verification of these findings by other therapists working with a less homogeneous group of clients must precede their acceptance.

About forty conclusions are stated by the authors, of which the following are examples: (1) It is possible for therapists to be nondirective in 98 per cent of their statements and have clients improve. (2) It may be possible to predict in early sessions which clients will or will not benefit from the nondirective approach. (3) The frequency of positive statements increases in a manner similar to the increase which occurs in individual therapy, but negative statements increase during the early sessions, then decrease during later sessions. (4) The least-profited clients tended to treat nondirective statements by the therapists as "threats" which put them on the defensive. (5) The most-profited clients tended to be more nondirective in dealing with other clients from the first session on. (6) Amount of participation by clients did not correlate with amount of improvement.

There are several methodological and technical weaknesses in the study. First, much is made of comparison of most-profited and least-profited clients. The authors used subjective criteria for evaluating improvement and present inadequate evidence of validity of these criteria. However, many of the conclusions are based on the assumption that the criteria were valid. Second, the authors confused nondirective, directive, interpretative, and critical statements in their classification of client responses. For example, they classed approval, encouragement, and reassurance responses as nondirective, and opinion, counteropinion, request-

ing client to elaborate, persuasion, suggestion, advice, interpretation, ta and deflection as "neutral." This ve unfortunate classification may inval: in date several of their conclusions 6 their generalizations about nondirec it tive client statements actually include o some directive categories, and gener ir alizations about evaluative and inter c pretative statements actually refer p to negative criticisms. Third, there b is a dearth of original data in this F book. Only three or four quotations t of client statements are presented in l the entire book. Consequently, the g face validity of the authors' inter a pretations and classifications has to t be assumed or challenged by the g reader. Finally, they appear at times v to overgeneralize and to conclude f that differences are significant or no i significant without applying tests o e significance.

WILBUR S. GREGORY
The University of Redlands

SCHEIDLINGER, SAUL. *Psychoanalysis and group behavior.* New York: W. W. Norton & Co., 1952. Pp. xviii+245. $3.75.

This book is an attempt to sum marize the "orthodox Freudian" point of view regarding the dynamics of group phenomena. As the author notes, there have been no scientific researches which have attempted to test Freudian hypotheses in this area so the source materials for this book are the theoretical concepts and hy potheses of Freud, Slavson, Fenichel, Redl, Money-Kyrle, and others. The author merits special commendation for stressing the hypothetical nature of the principles presented and the need for research (he devoted these entire seventh chapter to makes points)—an emphasis which makes this book somewhat unique among psychoanalytic writings.

The first five chapters are a detailed but repetitious and wordy survey of psychoanalytic literature dealing with group phenomena. Chapter is a recapitulation of them, and in the author does an excellent job of stating the concepts concisely and in a manner that makes it easy to contrast or integrate them with non-psychoanalytic approaches to group behavior. Although he mentions the Freudian principle that the family is the prototype of all groups and that leaders are father-substitutes, he gives equal emphasis to conscious and unconscious emotions, the functions of identifications, regression in groups, libidinal and aggressive motivations of group behavior, the benefits to the individual of membership in groups, factors which enhance or endanger group cohesion, the roles of leaders, and group climate. Although many of these principles are neither original with, nor limited to, the psychoanalytic viewpoint, all of them must be included in any summary of the psychoanalytic viewpoint.

In the last three chapters the author attempts to show the need for applying psychoanalytic principles to sociology, education, and group therapy. Although his theories regarding sociological research and his philosophy of education are stimulating, they are by no means original with, nor dependent upon, psychoanalytic theory. The discussion of group psychotherapy is of interest because it presents the various analytic viewpoints regarding the dynamics of group psychotherapy.

WILBUR S. GREGORY
The University of Redlands

VERNON, PHILIP E. *The structure of human abilities.* New York: Wiley, 1951. Pp. 160. $2.75.

In little more than a decade the

number of studies attempting to analyze the structure of large test batteries by use of factor analysis methods has mushroomed tremendously. As an aftermath to such vigorous activity it was inevitable that someone should try to draw a major segment of this somewhat contradictory, fragmentary array of studies into a closely knit, coherent presentation. That responsibility for undertaking such a task should be assumed by a prominent British psychologist is indeed a fortunate event for the American psychologist, who is usually too busy trying to keep up with the research of his associates to pay much attention to activities abroad.

In attempting to accomplish this objective, Vernon has done an unusually impressive job of thoroughly covering research pertinent to an understanding of human abilities. Frequently this has involved the reworking and reinterpretation of the results of many studies in order to fit them into a consistent pattern. His book, though small, is well organized and meaty, reading very much like a critical review article in the *Psychological Bulletin*.

While trying to be objective Vernon makes it clear from the beginning that he holds closely to the hierarchical group-factor theory of mental organization first put forward by Burt. Spearman's g permeates everywhere as a general factor distinguishable from two main group factors, the verbal-numerical-educational (referred to as the $v{:}ed$ factor) and the practical-mechanical-spatial-physical (called the $k{:}m$ factor). If an analysis is sufficiently detailed, these two major group factors can be split into numerous minor group factors which are of little importance according to Vernon. Although a major portion

of his book is devoted to a discussion of such so-called minor group factors, the author believes that little can be done to improve vocational guidance and selection after taking into account g, $v:ed$, and $k:m$. This is in sharp contrast to the prevailing American approach of designing test batteries for measuring a number of relatively independent ability factors from which an individual's pattern of factor scores can be derived and used in vocational guidance.

Vernon takes American factor analysts to task on several counts. Although the British factorists in the past may have been overcautious in extracting too few factors from a correlation matrix, most Americans have been far too indulgent, often extracting a dozen factors from a single matrix. While this is no doubt true, it should be pointed out that insignificant factors are frequently dropped from the analysis at a later stage when the original centroid factors are rotated into a more meaningful structure.

The point of greatest disagreement, however, concerns the role to be assigned a g factor. While British writers make g as large as possible, positing group factors only when the first centroid residuals require it, Americans tend to introduce it as a second-order factor or else rotate it completely out of the picture. Vernon argues that since g is so much larger than all other factors put together in *unselected* populations it is foolish to belittle it. Because most British factorists deliberately maximize g by choosing a set of reference axes which is favorable to their preconceptions, it is not surprising to find that g explains most of the common variance in many test batteries. But this is largely a matter of choice and the same correlation matrix can

be explained just as adequately (and in the opinion of many factorists more meaningfully!) by a rotation of the reference frame which minimizes g and results in multiple-group factors.

Seven main arguments have been put forth in the Appendix to support the writer's stand that his hierarchical group-factor theory is superior to any other. While several good points are made and will prove stimulating to the reader, it is doubtful that they are sufficiently crucial and persuasive to convince most factor analysts who have found the multiple-factor methods of Thurstone satisfactory.

This is the clearest, most consistent, and well-documented presentation to date of the hierarchical group-factor theory which dominates the British scene. Regardless of one's personal predilection concerning factor analytic procedures, careful study of this little book will prove highly stimulating and informative.

WAYNE H. HOLTZMAN
University of Texas

HIRSH, IRA J. *The measurement of hearing.* New York: McGraw-Hill 1952. Pp. ix+364. $6.00.

The purpose of this book is to provide background material in acoustics, electronics, psychophysics, and psychoacoustics for those who are entrusted with clinical audiometry. No previous technical knowledge is required to read it, the matter of clear nonmathematical definitions being given especial care. It is excellent in the initial training of individuals who must be assigned to clinical audiometry but who bring no background in the psychology of hearing and have had no experience with the psychophysical method. It has more limited value for those who can be asked to read chapters on sound and elementary electronics in physics

exts, Guilford's *Psychometric Meth-ds*, and the chapters on acuity, dif-erential sensitivity, and loudness in *Hearing* by Stevens and Davis.

The book is not, however, a simple manual in the techniques of audi-ometry. After perusing it, the reader would not at all know how to go about taking an audiogram. This is not bad—there are are plenty of man-uals and texts on that topic, though standardization is yet to come. But it emphasizes the fact that this par-ticular book is a serious attempt to examine the psychological back-grounds of collecting absolute inten-sive limens, and the psychological, not the clinical, meaning of those limens once collected. No other source does just this, and the book very definitely fills a need. It may surprise a psychological audience that their contributions to the psycho-physics of threshold determination and to sensory scaling have not infil-trated more widely into the new pro-fession of audiology. But the re-viewer has lectured to graduate stu-dents in that field none of whom had thought of defining their stimulus in physical terms of sound pressure level in a closed coupler or in the ear canal itself; they were content to accept the hearing loss dial-reading on the audiometer. And there have been applications for employment from doctorates in audiology who could not attempt a definition of the method of limits. It is clear that the profession of audiology would be much better grounded in audiometry, at least, if this book were required reading.

As a reference text in the psychol-ogy of hearing, the book has some additional values. There is an excel-lent chapter on loudness and the recruitment of loudness, one on binaural hearing and bone conduc-

tion, and one on masking and fatigue. These, quite the best chapters as reference, give a good deal of up-to-date primary source material. No more errors are committed than are unavoidable in making generaliza-tions from research material for clini-cal use. The book is carefully built with ample subject and name indexes, glossary, and appendices of audi-ometer specifications and several word lists for speech reception test-ing. It is a pleasure to recommend this book to audiometricians, and to psychologists doing research in hear-ing who want to explore the relation of their work to the clinic.

J. DONALD HARRIS
U. S. Naval Medical Research Lab-oratory, New London, Conn.

ASHBY, W. ROSS. *Design for a brain.* New York: Wiley, 1952. Pp. ix+259. $6.00.

Perhaps the future historian will label the psychologists of today the "model builders." In any event, we are experiencing great efforts to create models that may prove more fruitful in understanding behavior than our more mundane empiricism of the past. Ashby's book is another of these current attempts to develop a model of the brain as the principal organ of behavior.

Ashby, like the others, uses the machine as his model but his model has the merit of being a generalized machine, not a servo-mechanism nor a computer. Indeed, Ashby's model is really a mathematical one. He conceives the brain to be a system of variables with certain functional properties. He only uses a machine to illustrate these properties, not as the analogue of the brain.

To understand the model, one has to read the book carefully and digest each of many steps that are essential

to his argument. Ashby presents them in nonmathematical form in his main text, leaving the proof to an appendix. The skeleton of the argument is this: The brain is a dynamic system of variables, a system in which values of variables are constantly changing. At any one moment these values constitute a state, and the system moves from one state to another. The path established by successive states is the *line of behavior*.

To explain behavior we must assume the system to have certain properties. For one thing, the system is *absolute;* given any one state, all lines of behavior following that state are equal—it is a completely state-determined system. It is also a *stable* system, for its variables are so interrelated that they cannot exceed certain limits; it has feedback in it that makes it a homeostatic system. It is a system which at times enters *critical states* that produce relatively sudden changes in certain of its variables, so that its variables describe *step functions;* like a thermostat, one variable suddenly changes when another reaches a certain value. Critical states and step functions combine to endow the system with *ultrastability*, that is, with preferences for states that do not lead to critical states. Finally, many of the variables in the system are *part functions* of other variables; that is, they are sometimes dependent on, and sometimes independent of, each other.

Assuming these and other mathematical properties of the brain, Ashby attempts rather successfully to deal with a wide variety of phenomena including physiological homeostasis and survival, adaptation, learning, and goal-seeking. He even seems to prove that learning and memory must not be localized very well in the nervous system but rather dispersed,

as one might conclude from Lashley's work.

This reviewer is not at all confident of the outcome of all of this model building. If it is a good thing, he prefers Ashby's formal mathematical approach to the mechanical and electrical analogies of such model builders as McCulloch and Wiener. Ashby, moreover, has worked out a general, comprehensive model, in contrast to the piecemeal approach of many model builders. His key concept, that of ultrastability, is a feature not existing, or at least not well developed, in other models and it seems to have potential power to explain the knottier problems of motivation, purpose, and complex mental functioning. The book, therefore, contributes importantly to a mathematical formulation of behavioral phenomena.

C. T. MORGAN
The Johns Hopkins University

THURSTONE, L. L. (Ed.) *Applications of psychology.* New York: Harper, 1952. Pp. x+209. $3.00.

Any book which is the joint product of several authors will necessarily display the diversity of viewpoint and styling which are characteristic of those who have participated. *Applications of Psychology* is a good example.

Prepared in honor of the seventieth birthday of Walter V. Bingham, the contributors were chosen from among the many who had been associated with Dr. Bingham either at Carnegie Institute of Technology or during his two periods of service as a military psychologist. These are representative of the many current areas in applied psychology.

The topics covered range from somewhat theoretical discussions of creative talent through the validity

f the medical and life insurance
cales of the Strong Vocational Inter-
st Blank to the clinical evaluation
f Harvard undergraduates and the
se of the clinical method in the selec-
ion of employees.

The importance of the volume
would seem to lie chiefly in its tracing
of the variety of present-day prac-
ices to their early beginnings and in
ts documentation of the effects of
Bingham's influence on the develop-
ment of many phases of professional
psychology.

As a testimonial to the respect and
admiration accorded to Bingham, it
is an effective document. As a com-
munication in the field of psychology,
it leaves much to be desired.

GEORGE K. BENNETT

The Psychological Corporation

WOLFF, W. *The dream—mirror of
conscience.* New York: Grune &
Stratton, 1952. Pp. vi+348. $8.50.

The sense of diffuse dissatisfaction
with the first section of this book
(dream and history) is perhaps the
consequence of its inevitably episodic
character; inevitably, because the
author could naturally hit in the
most fleeting fashion only the highest
of the high spots in presenting Greek,
Roman, and Medieval treatments
and views of the dream as well as
Egyptian, Babylonian, and Biblical
accounts. With chapters running
about a page and a half to two pages
one can hardly be left with any other
feeling than that something is un-
finished. However, any other treat-
ment would have required a much
larger book and perhaps have
changed its slant and intent.

With respect to a few minor dis-
agreements, in the first place it might
be that the author takes too literally
Freud's complaint that "nothing new
or valuable" has been written about

the dream since 1900. One might
grant the stamp of genius almost to
all of Freud's works and yet recognize
that for him the "value" of a theoreti-
cal position was measured, to some
extent, by the degree to which it was
in agreement with his own. Whether
one characterizes Stekel's analysis of
dreams as "new" or "valuable," it
was certainly *different.* My impres-
sion, too, is that Wolff emphasizes too
little the distinction that since has
been made between Freud's view
that the dream tapped the "archaic"
in the personality and the view that
the "higher" facets function in the
dream. The latter case has been
strongly argued by Fromm, whose
book is not listed in Wolff's. Thus,
Wolff's distinction between "the
forces that drive man" and "the man
using these forces" (p. 280) is cer-
tainly not a new one. Indeed, unless
I am in error, this differentiation is
expressed in the distinction between
the "old" or earlier psychoanalysis
and the more recent "ego psychol-
ogy."

Not withstanding these minor criti-
cisms, Wolff's book is stimulating,
provocative, and rewarding. There
were two outstandingly significant
notions. The "theoretical" one is
perhaps best expressed in Wolff's
own words: "the dreamer uses his
dream activity to solve this conflict
of needs which conscious reasoning
could not solve because the conflict
usually involves antagonism between
emotion and reason. Thus, a third
agency is needed to decide this
struggle. This is conscience . . . the
dreamer . . . confronts his thoughts
with his values and brings them be-
fore the mirror of his conscience" (p.
275).

The second major contribution is
a proposal for a method of dream
analysis or "synthesis." Criticizing

orthodox analytic procedure, in that "psychoanalysts only pick out those dream images which they consider the most significant" (p. 189), Wolff proposes that the psychotherapist "should use *all* the dream images" (p. 190). Further, "we must eliminate not only our personal preferences and leading ideas in a dream interpretation but also those of the dreamer himself. If we give him the dream images as stimuli for his associations in the sequence of the dream, he may pattern his associations according to this sequence. In order to prevent such a preconceived arrangement we give the dream images as stimuli in a sequence different from that of the dream. Moreover, in order to free his associations, the dream stimuli should be given a few days after the dream was told in such a way that the dreamer will not necessarily recognize the words as coming from his dream" (p. 190). For the experimental-minded clinician, this proposal is worthy of investigation.

WILLIAM SEEMAN

Mayo Clinic
Rochester, Minn.

WHITE, ROBERT W. *Lives in progress*. New York: Dryden Press, 1952. Pp. ix+376. $2.90.

This is an important book. It is a valuable addition to the slowly growing shelf of books that provide raw material of psychology suitable for undergraduate students. The book will be welcomed by psychology teachers who yearn for freedom from the tyranny of textbooks. Approximately 50 per cent of *Lives In Progress* is devoted to the life histories of three young adults whom White and his associates studied intensively when the subjects were college students and studied again five to ten years later. The remainder of the book is an exposition of methods of personality study and an analysis of these lives in terms of White's views of personality development. In the reviewer's opinion, the analysis provides an unusually adequate basis for teaching current theories of personality to undergraduate students; its chief virtue lies in its comprehensive and balanced coverage. *Lives In Progress* provides a basis upon which an instructor can develop his own brand of personality theory. Since the case histories are presented with sufficient detail and absence of bias, they can be used to exemplify other viewpoints and emphases than those of the author.

The most obvious use of this book is in undergraduate courses in personality. However, the reviewer has used it very successfully in general introductory psychology, along with a volume presenting another kind of unanalyzed psychological data (Barker and Wright's *One Boy's Day*). These two books freed the instructor to be a teacher of psychology rather than a textbook commentator. *Lives In Progress* has high interest for students; the reviewer found that some students read the book in one sitting and many did so during the first week of the course. This attraction is apparently due partly to the intrinsic interest of the life histories, but also to the fact that White is an excellent writer.

Although the author apparently considers *Lives In Progress* to be largely a teaching resource, it is the reviewer's opinion that it, along with such volumes as Blos' *The Adolescent Personality* and Davis' *Children in Bondage*, is a significant contribution to the small stack of normal lives available in published form to students of personality. The scientific value of these case studies will not

diminish with the passage of time. They may indeed turn out to be important scientific documents a century hence. One can easily understand the historical and psychological interest an 18th or 17th century *Lives In Progress* would have today.

ROGER G. BARKER
University of Kansas

FERGUSON, LEONARD W. *Personality measurement*. New York: McGraw-Hill, 1952. Pp. xiii+457. $6.00.

A volume devoted to the special problems of personality measurement has been needed for a long time, but Ferguson fills only a portion of this gap. A definition of personality is lacking, which would serve as a focal point for the measurement material. Many problems in the logic of measurement are sidestepped, such as the extent to which the addition of item-responses on interest inventories, questionnaires, and projective devices gives us scalable material. Ferguson's book, therefore, must be thought of primarily as a collection of interesting and useful facts about the reliabilities and validities of numerous instruments intended to be measures of aspects of personality.

The first chapter of the book is well calculated to stir the reader's enthusiasm. A case of a college student suicide suggests that we may get interesting clinical studies on validity; a discussion of group differences in personality (national, racial, etc.) hints of measurement problems in cross-cultural comparisons; and three pages on morale whet an appetite for a critical analysis of morale measures. Unfortunately, most of these expectations are unfulfilled; cultural group differences in personality are not mentioned again, and morale studies are given only casual attention. There is a good deal to be said in

favor of Ferguson's arrangement, starting with problems of interest and attitude measurement, which makes it possible to deal with simple concrete materials before getting into some of the more complex devices. It is questionable, however, whether these two topics merit 116 pages (or 25 per cent of the volume) compared to the treatment given to projective devices in 35 pages (or 8 per cent of the book). The suggested imbalance in Ferguson's presentation is further illustrated in the treatment of the projective tests. The chapter on the Rorschach test gives a protocol of a fairly interesting case, and discusses some of the mechanical aspects of scoring. However, at no point is the interpretation quoted; that is, the description of the total personality, which is presumably the main diagnostic claim of the Rorschach approach. Even on the level treated by Ferguson (the statistical analysis of specific response determinants), the work of such investigators as Wittenborn and Cronbach has been ignored. And just why the study of sales managers by Kurtz should be considered evidence on the validity of the Rorschach is hard to see.

The sections on objective, questionnaire-type devices are the best in the book. They summarize data on an enormous number of different inventories, from the Woodworth Personal Data Sheet to the Minnesota Multiphasic Personality Inventory. And information on method of development, reliability, group norms, and validity studies is offered for each of these tests. There is lacking, however, a discussion of such logical problems as what kind of scale results when we simply add the number of "diagnostic" answers. What one also misses is an incisive discussion of just what would constitute

substantial validation. This is a difficult assignment, but it cannot be evaded legitimately in a volume on personality measurement.

This book should be useful as a text for a graduate course in personality measurement, if it is liberally supplemented by journal articles or other sources, and if the instructor provides some balance to the overemphasis on empirical findings and on questionnaire-type tests. The treatment is clear and straightforward; the style is readable without being overpopularized. It will certainly help to give students a model of organization for data on development, reliability, and validity of any test device. They should also learn a good deal about appropriate applications of statistics to defined measurement problems.

Ross Stagner
University of Illinois

Vernier, Claire Myers. *Projective drawings.* New York: Grune & Stratton, 1952. Pp. v+168. $6.00.

This is a well-organized collection of drawings made by persons of a variety of clinical diagnostic groups (the psychoses, neuroses, and brain-damaged) and includes the productions of normal subjects as well. The author offers this book primarily as a teaching tool in courses on projective techniques, and for this end it should be most valuable. Cases were selected for which the clinical diagnoses were well defined and the drawings were then collected to see the extent to which they reflected the

already known symptoms. In surveying these productions it is certainly evident that the "Draw-A-Man" test can be valuable in capturing graphically some personality characteristics. What is also strikingly evident, however, is the tremendous overlap of so-called clinical signs from one diagnostic group to another, especially into the normal group. This should allow for considerable humility and care in the use and interpretation of the test until a more standardizing system of dealing with the data can be developed. Vernier data may present a significant step in this direction. It is important to recognize, however, that most of the observations made about the drawings by the author are ex post facto which sometimes get her into trouble. In one instance drawings made by a paranoid patient (Figure 2), in which the figures both had small heads, were interpreted as follows: "minimal head emphasis is of interest in view of the patient's lack of either intellectual aspirations or control mechanisms" (p. 6). However, in a drawing made by another paranoid schizophrenic (Figure 4) there is also a small head which is not mentioned; instead, the midline emphasis of a double belt line is considered to be a reflection of the patient's emphasis on intellectual controls. There are a large number of similar examples. In general, however, this book should be of considerable value for the teacher of projective techniques, for the practicing clinician, and in suggesting further research with this test.

Irene R. Pierce
Wellesley College

American pocket medical dictionary. (19th Ed.) Philadelphia: W. B. Saunders, 1953. Pp. iv+639.

ARNY, CLARA B., DYER, DOROTHY T., & PROSHEK, MARGARET F. *Minnesota tests for household skills.* Chicago: Science Research Associates, 1952. Pp. 7 (manual), 5–7 (tests). $.25 (manual), $.75 (tests).

ASHER, ESTON J., TIFFIN, JOSEPH, & KNIGHT, FREDERICK B. *Introduction to general psychology.* Boston: D. C. Heath, 1953. Pp. xvi+515. $4.25.

BAKER, HARRY J. *Introduction to exceptional children.* New York: Macmillan, 1953. Pp. xvi+500. $5.00.

BERNHARDT, KARL S. *Practical psychology.* (2nd Ed.) New York: McGraw-Hill, 1953. Pp. xii+337. $3.75.

BLUM, GERALD S. *Psychoanalytic theories of personality.* New York: McGraw-Hill, 1953. Pp. xviii +219. $3.75.

BONNER, HUBERT. *Social psychology: an interdisciplinary approach.* New York: American Book Co., 1953. Pp. 439. $4.25.

COMMITTEE ON COLORIMETRY OF THE OPTICAL SOCIETY OF AMERICA. *The science of color.* New York: Thomas Y. Crowell, 1953. Pp. xiii+385. $7.00.

CRUZE, WENDELL W. *Adolescent psychology and development.* New York: Ronald Press, 1953. Pp. xii+577.

FLEW, A. G. N. (Ed.) *Logic and language.* New York: Philosophical Library, 1953. Pp. vii+242. $4.75.

GELLHORN, ERNST. *Physiological foundations of neurology and psychiatry.* Minneapolis: Univer. of Minnesota Press, 1953. Pp. xiii +556. $8.50.

GRASSI, JOSEPH R. *The Grassi block substitution test for measuring organic brain pathology.* Springfield, Ill.: Charles C Thomas, 1953. Pp. ix+75. $3.00.

HALMOS, PAUL. *Solitude and privacy.* New York: Philosophical Library, 1953. Pp. xvii+181. $4.75.

HILGARD, ERNEST R. *Introduction to psychology.* New York: Harcourt, Brace, 1953. Pp. x+192. $5.75.

KATZ, BARNEY, & LEHNER, GEORGE F. J. *Mental hygiene in modern living.* New York: Ronald Press, 1953. Pp. x+544. $4.50.

KROEBER, A. L., & KLUCKHOHN, CLYDE. *Culture: a critical review of concepts and definitions.* Papers of the Peabody Museum of American Archaeology and Ethnology, Harvard Univer. Cambridge, Mass.: Peabody Museum, 1952. Pp. viii+223. $5.25.

LANSING, ALBERT I. (Ed.) *Cowdry's Problems of ageing: biological and medical aspects.* (3rd Ed.) Baltimore: Williams & Wilkins, 1952. Pp. xxiii+1061. $15.00.

LAWSHE, C. H. *Psychology of industrial relations.* New York: McGraw-Hill, 1953. Pp. vii+350. $5.50.

LUNDIN, ROBERT W. *An objective psychology of music.* New York: Ronald Press, 1953. Pp. ix+303. $4.50.

MACE, C. A., & VERNON, P. E. *Current trends in British psychology.* London: Methuen, 1953. Pp. viii+262. 15s. net.

MINER, ROY W. (Ed.) *Comparative conditioned neuroses.* Annals of the New York Academy of Sciences, Vol. 56, Art. 2. New York:

Academy of Sciences, 1953. Pp. 141–380. $3.50.

PATTERSON, C. H. *The Wechsler-Bellevue scales: a guide for counselors.* Springfield, Ill.: Charles C Thomas, 1953. Pp. viii+146. $3.75.

POWDERMAKER, FLORENCE B., & FRANK, JEROME D. *Group psychotherapy: studies in methodology of research and therapy.* Report of a group psychotherapy research project of the U. S. Veterans Administration. Cambridge, Mass.: Harvard Univer. Press, 1953. Pp. xv+615. $6.50.

RANSON, STEPHEN W., & CLARK, SAM L. *The anatomy of the nervous system.* (9th Ed.) Philadelphia: W. B. Saunders, 1953. Pp. xii +581. $8.50.

REISS, SAMUEL. *The universe of meaning.* New York: Philosophical Library, 1953. Pp. x+227. $3.75.

RIESE, WALTHER. *The conception of disease: its history, its versions and its nature.* New York: Philosophical Library, 1953. Pp. 120. $3.75.

RUNES, D. D. *The Soviet impact on society.* New York: Philosophical Library, 1953. Pp. xiii+202 $3.75.

SENN, MILTON J. E. (Ed.) *Problems of infancy and childhood.* Transactions of the Sixth Conference Josiah Macy, Jr. Foundation. New York: Josiah Macy, Jr. Found. 1953. Pp. 160. $2.50.

TRAXLER, ARTHUR E., JACOBS, ROBERT, SELOVER, MARGARET, & TOWNSEND, AGATHA. *Introduction to testing and the use of test results in public schools.* New York: Harper, 1953. Pp. x+113. $2.50.

TUCKMAN, JACOB, & LORGE, IRVING *Retirement and the industrial worker* New York: Bureau of Publications Teachers Coll., Columbia Univer. 1953. Pp. xvi+105. $2.75.

WALTER, W. GREY. *The living brain.* New York: W. W. Norton, 1953. Pp. 311. $3.95.

WESCHLER, IRVING R., & BROWN, PAULA. (Eds.) *Evaluating research and development.* Los Angeles: Human Relations Research Group, Institute of Industrial Relations, 1953. Pp. 104. $1.65.

Psychological Bulletin

A BRIEF HISTORY OF CLINICAL PSYCHOLOGY

ROBERT I. WATSON
Northwestern University[1]

Clinical psychologists have been surprisingly ahistorical. Very little thought has been given to, and less written about, the origin and development of clinical psychology. In the literature there are articles and books which interpret historically various special aspects or evaluate related fields, some of which have been of considerable help in preparing this paper. Nevertheless, whatever the reasons, there is no available general account of the history of clinical psychology from the perspective of today.

In part, this neglect is due to the upsurge of interest in clinical psychological activity during and following the second World War. Since then, clinical psychologists have had little time to spend inquiring into their origins. Then too, their day-to-day activities impress them as so new and vital that they are hardly to blame for tacitly accepting the belief that they are pioneers and that somewhere in the chaos of war and its aftermath was born a new profession having little or no relation to what went before. The state of affairs today is curiously reminiscent of the situation found by Kimball Young in 1923 in tracing the history of mental testing. He remarked, "Making history on

[1] This article was written while the author was at the Washington University School of Medicine.

every hand as we are, we have a notion that we somehow have escaped history" (121, p. 1).

To capture in full measure the sweep and continuity of the history of clinical psychology is beyond the competence of the reviewer, to say nothing of space limitations. In order to do justice to all aspects of the subject one would have to deal with the complex history of the psychology of motivation and dynamic psychology. Similarly, all the ramifications of the relation of clinical psychology to the rest of the field of psychology, of which it is an integral part, as well as an account of the history of test development would have to be considered.

The present account, perforce, presents an examination of men and ideas influential in shaping clinical psychology. But, since psychology is now a profession, attention must also be devoted to those internal and external controls which characterize a profession and to the settings in which the professional practice is conducted.

In presenting a historical account the question arises concerning the most appropriate date at which to begin. With some justification it was decided that since clinical psychology, as we know it, arose at about the turn of the present century it would be appropriate to begin with the im-

mediate forerunners of this first generation of clinical psychologists. The origins of clinical psychology, the first major section of this account, are to be found in the psychometric and dynamic traditions in psychology; the psychologist in the settings of the psychological clinic, child guidance, mental hospitals, institutions for the mentally defective; and the beginnings of psychology as a profession. Somewhat arbitrarily this early pioneer work is considered to come to a close with the end of the second decade of this century. This is followed by a section concerned with clinical psychology in the twenties and thirties. The same topics just mentioned, e.g., the dynamic tradition and psychology as a profession, are again considered. The work of psychologists in the armed services during the second World War and its effect upon psychology in the postwar period are next evaluated. A brief overview of clinical psychology today closes the account.

THE ORIGINS OF CLINICAL PSYCHOLOGY

The Psychometric Tradition in Psychology

This tradition, one of the headwaters from which clinical psychology sprang, was, in turn, a part of the scientific tradition of the nineteenth century. With all the limitations with which it is charged today, it is to this movement that the clinical psychologist owes much of his scientific standing and tradition. Whenever a clinical psychologist insists upon objectivity and the need for further research, he is, wittingly or otherwise, showing the influence of this tradition. Moving from Galton through Binet and Terman, this tradition met the demand that psychology, if it was to become a science, must share with other sciences the respect for quantitative measurement.

Psychometrics as a tool for clinical psychology owes its beginnings to Francis Galton (53) in England. Grappling as he was with the problem of individual differences, he and his followers did much to lay the groundwork for the investigation of ability by using observations of an individual's performance for information on individual differences. He thus founded mental tests.

In 1890 Cattell (29) introduced the term "mental tests" in an article describing tests which he had used at the University of Pennsylvania. Even at that date he was pleading for standardization of procedure and the establishment of norms. From the time of his days as a student of Wundt's, Cattell was interested in the problem of individual differences and did much to stimulate further investigation. Along with Thorndike and Woodworth, he also stressed dealing with individual differences by means of statistical analysis—a really new approach at this time. Some of these investigations, both from Cattell's laboratory and from others in various parts of the country, made positive contributions to various facets of the problem of psychometric measurement. For example, Norsworthy (82) in 1906 compared normal and defective children by means of tests and found the latter not a "species apart," pointing out that the more intelligent of the feebleminded were practically indistinguishable from the least intelligent of the normal.

Most of the investigations of the time were concerned with simple sensorimotor and associative functions and were based on the assumption that intelligence could be reduced to sensations and motor speed,

an attempt which, as is now known, was doomed to failure. Furthermore, although more suitable verbal material was used, the studies of college students at Cornell, such as Sharp's (95), and the Wissler study (113) at Columbia, were found to be essentially nonpredictive. What the workers failed to take into account was the fact that college students are a highly selected group having a considerably restricted range. The negative finding of these studies effectively blocked further investigation at the college level for years. When one stops to consider that the dominant systematic position of the day was the structuralism of Titchener, who had banished tests as nonscientific, it is no wonder that "tests" were viewed with at least a touch of condescension.

In the meantime Binet had been working in France developing his tests based on a wider sampling of behavior than had yet been used. His success in dealing with the intellectual classification of Paris school children is well known and needs no amplification at this point. The translation of his tests and their use in this country followed shortly after the turn of the century. It was Goddard (simultaneously with Healy), a student of G. Stanley Hall, who introduced the Binet tests to this country. Through a visit abroad and contact with Decroly, he became acquainted with Binet's work (121). In 1910 he began publishing findings with the test and in 1911 published his revision of the 1908 Binet Scale. This revision, along with Kuhlmann's, also published in 1911, gained some popularity among clinicians, but the subsequent development by Terman far overshadowed their work.

Probably the test that had the most influence upon trends in clinical psychology was the Terman Revision of the Binet Scale (83). In fact, for years the major task of the clinical psychologist was to administer the Stanford-Binet. In view of the importance of this test it is desirable to present in some detail the background of its development.

Lewis M. Terman (101) received his graduate training at Clark just after the turn of the century under Hall, Sanford, and Burnham. As Terman put it, "For me, Clark University meant briefly three things: freedom to work as I pleased, unlimited library facilities and Hall's Monday evening seminar. Anyone of these outweighed all the lectures I attended" (101, p. 315). This influence of Hall's was more from the enthusiasm he inspired and the wide scope of his interests than from his scientific caution and objectivity. Sanford was his doctoral adviser, but Terman chose his own problem in the area of differentiation of "bright" and "dull" groups by means of tests and worked it through more or less independently.

By a severely limited survey such as this it would be easy to give the impression that little or nothing else was being done along the lines under discussion except that reported. Terman was not alone in his interest in the development and standardization of tests by any manner of means. In his autobiography Terman (101) mentions as known to him in 1904 the work of Binet, Galton, Bourdon, Oehrn, Ebbinghaus, Kraepelin, Aschaffenburg, Stern, Cattell, Wissler, Thorndike, Gilbert, Jastrow, Bolton, Thompson, Spearman, Sharp, and Kuhlmann.

At the suggestion of Huey, who had been working in Adolf Meyer's clinic at Johns Hopkins, Terman, undeterred by the prevailing hostile attitude of most psychologists, began

work with the 1908 Binet Scale and in 1916 published the Stanford Revision of the Binet-Simon tests. Terman's interest in both the test and results from it continued unabated, resulting in still another revision in 1937.

Performance tests, so necessary for work with the linguistically handicapped, actually antedated the Stanford-Binet. The Seguin, Witmer, and Healy form boards and other performance tests were already in clinical use. Norms, although not lacking, were undeveloped, and the directions placed a high premium on language. What seemed to be needed was a battery of performance tests sampling a variety of functions and not as dependent upon language. Among the earliest to appear and to come into fairly common use was the Pintner-Paterson Scale of Performance Tests (85), published in 1917. Included in this scale were several form boards, a manikin and a feature-profile construction test, a picture completion test, a substitution test, and a cube-tapping test.

Another major step was the development of group tests under the impetus of the need for large scale testing of recruits in World War I. This testing program is described with a wealth of detail by Yerkes (120). Although group tests were not unknown before the war, as witness those described in Whipple's *Manual of Mental and Physical Tests* (110), the need for quick appraisal of the basic intelligence of a large number of men provided the impetus for extensive development. The *Alpha* scale for literate English-speaking recruits and the *Beta* scale for illiterates and non-English-speaking recruits were developed rapidly under this demand. The Woodworth Personal Data Sheet (118), the first of a long line of psychoneurotic inventories, also was a product of military needs. So successful were these tests in overcoming the prejudices against testing both within the field of psychology and in the general public that after the war a veritable flood of group tests appeared. Many extensive surveys in the public schools were made for classificatory purposes. Further developments in this tradition during the twenties and thirties will be appraised after examination of other aspects of the origin of clinical psychology.

The Dynamic Tradition in Psychology

A major source of influence contributing to the growth of clinical psychology was the thinking and writing of the "Boston group" who promulgated "the new psychology" —William James, G. Stanley Hall, and their associates. Although in no way could they be labeled clinical psychologists, their thinking was much closer to the heart of the clinical psychology movement and to progressive psychiatry than was the structural point of view of Titchener. Heresy though it may be, it cannot be denied that at that time academic psychology had relatively little to contribute to clinical psychology. Psychology, to be sure, had been placed by Fechner, Helmholtz, Wundt, Kraepelin, and others upon a scientific, quantitative foundation instead of being permitted to remain an indistinguishable cohort of philosophy. This was an essential step without which there could have been no clinical psychology; nevertheless, a sensationalistic approach to conscious intellectual experience offered relatively little for the clinical method and the profession with which it was to be associated.

The psychiatry of the day was in

the main concerned with pathology and the search for an explanation of mental disturbances in disease processes. Kraepelin (68) had introduced clarity through his classification of mental disease, but at the expense of deeper understanding. Based upon symptoms and primarily descriptive in character, his classification served to diminish—even to discourage—in its users any urge toward understanding of psychological dynamics.

French psychiatric thinking and research profoundly influenced James (80). The work of Janet and Charcot was particularly important in this connection. With Morton Prince, he did much to stimulate interest in the phenomena of dissociation, feeling as he did that it was a fruitful method of investigation of personality functioning. Early in his career he recognized the value of a clinical approach which led him "whenever possible to approach the mind by way of its pathology" (77, p. 20).

The influence of James was expressed primarily through his *Principles of Psychology* (65), published in 1890, and to a lesser degree by his *Varieties of Religious Experience* (66), published in 1902. Both of these works were sufficiently removed from the otherwise prevailing psychological thinking of his day to be considered major pre-Freudian, dynamic influences. The choice of the term "dynamic" in this context is neither idle nor wishful thinking. James himself used the term to distinguish his point of view from the structural approach of Titchener (86).

Concerning the influence of the *Principles*, Morris had this to say:

Great books are either reservoirs or watersheds. They sum up and transmit the antecedent past, or they initiate the flow of the future. Sixty years after its publication, the *Principles* appears to be one of the major watersheds of twentieth-century thought. Directly or indirectly, its influence had penetrated politics, jurisprudence, sociology, education and the arts. In the domain of psychology, it had foreshadowed nearly all subsequent developments of primary importance. Viewed retrospectively, the permanent significance of the *Principles* was incentive. It explored possibilities and indicated directions. These led, eventually, into social, applied and experimental psychology; into the study of exceptional mental states, subliminal consciousness and psychopathology. Because of its extreme fertility in the materials for hypothesis, most of the competitive schools of psychological theory that arose during the first half of the century could claim common ancestry in the *Principles* for at some point it implied their basic assumptions (77, p. 15).

This aptly catches James's influence on clinical psychology, not through work directly in the field or with the method, but through the fertile (and contradictory) character of his thinking.

In addition to the stimulation of his writings, James did take specific action of direct relevance to clinical psychology in his support of Clifford W. Beers, whose book, *A Mind that Found Itself* (19), did so much to further the mental hygiene movement. This he did through an endorsing letter which appeared in the first edition and, according to Henry James, his son, by departing from his fixed policy of "keeping out of Committees and Societies" (64, p. 273). In addition, he was interested in psychical research and in the efforts of Freud and Jung, although dubious about both of these trends (64).

Obviously it is impossible to capture the full flavor of William James in a paragraph or two, but this "defender of unregimented ideas" is at least the eccentric brilliant uncle of

the men in clinical psychology who followed after.

Another of the pioneers of this time and place was G. Stanley Hall. He was more influenced by the evolutionary concept stemming from Darwin than by French psychopathological thinking. Shakow, in considering Hall's influence on psychiatry, so well summarizes his contributions that they may be seen as contributions to clinical psychology as well. He writes that it was:

Hall, the propagandist, who gave Freud his first academic hearing, who gave courses in Freudian psychology beginning in 1908 and whose pressure for its consideration remained life-long; Hall, who influenced Cowles in establishing the psychological laboratory at McLean Hospital which had as directors following Hoch, Franz, Wells and Lundholm; Hall, who stimulated Adolf Meyer, by his early interest in child study, to write his first paper on a psychiatric topic— *Mental Abnormalities in Children during Primary Education . . .*—and who did so much to make the country child-conscious; Hall, whose students Goddard and Huey (also Meyer's students at the Worcester State Hospital) did the early pioneer work on feeblemindedness . . . Hall, whose bravery in handling the problem of sex did so much to break down the first barriers, thus greatly facilitating the later child guidance handling of this and related problems; Hall, whose student Terman achieved so much in the development of the Binet method in the United States and whose student Gesell did so much for other aspects of developmental psychology; Hall, whose journals regularly published material of psychopathological interest; Hall, the ramifications of whose psychological influence are most pervasive in fields related to psychopathology . . . (92, p. 430).

Certain other factors might also be mentioned. Before his period as president of Clark University, Hall, while at Johns Hopkins, held weekly clinics at Bay View Hospital and, until its medical staff was organized under his direction, served as lay superintendent. For a period of years he taught and demonstrated for psychiatrists at Worcester State Hospital, handing over the actual instruction in 1895 to Adolf Meyer, but continuing his interest in the field (74). Other students of this period who made substantial contributions to clinical psychology included Blanchard, Conklin, Kuhlmann, and Mateer.

Something of the spirit and activity of the associates of these men may be captured by an examination of the journal that was begun early in the century. *The Journal of Abnormal Psychology*, later called *The Journal of Abnormal and Social Psychology*, was a major source of publication of the more enlightened efforts of its time. Until 1913, when the *Psychoanalytic Review* was founded, it was the only journal in which psychoanalytic papers were published (32). Founded in 1906 for the express purpose of serving both medicine and psychology, it had as its editor Morton Prince, later professor of psychology at Harvard University, and numbered among its associate editors Hugo Münsterberg, James Putnam, August Hoch, Boris Sidis, Charles L. Dana, and Adolf Meyer. The papers in the first issue aptly catch the various influences at work in the psychology and psychiatry of the day. The first is a paper by Janet and thus represents the French psychopathological school; the second concerns hypnosis; the third, a critique of Freud by Putnam (the first article in English calling attention to Freud's work); and the fourth, a paper by Morton Prince concerning his most famous case of multiple personality, Miss Beauchamps. The

first book review in this new journal was that of Freud's *Psychopathology of Everyday Life*, which had been published in Germany in 1904. So far as this writer is aware, the first critical article concerning psychoanalysis by an American psychologist appeared in the February 1909 issue of this journal. It was entitled "An Interpretation of the Psychoanalytic Method in Psychotherapy with a Report of a Case so Treated" (90). This is apparently the second instance of a report of personal psychotherapeutic experience by a psychologist.[2] Its author, known for endeavors in fields far removed from this, was Walter Dill Scott, the psychologist, later president of Northwestern University.

The situation in the official psychiatric journal may be used for contrast. The first psychoanalytic paper to appear in the *American Journal of Insanity* was in the October 1909 issue. This paper was by Ernest Jones of Toronto and deplored the fact that Freud's methods had been neglected. None of Freud's books was reviewed in this journal for some years and, indeed, the first review to appear was that of Brill's *Psychoanalysis* in July 1914.

Isador Coriat (32), in presenting some reminiscences of psychoanalysis in Boston, attributes the interest in psychotherapy there to the stimulation of William James. Although A. A. Brill began psychoanalytic practice in New York in 1908, he was the only psychiatrist in the United States at that time engaging in such practice. He, Putnam, and Ernest Jones,

then of Toronto, were the first in America to do active work with psychoanalytic methods. The first English translation of a work by Freud, *Selected Papers on Hysteria*, appeared in 1909 according to Coriat (32). It was in this same year that G. Stanley Hall, as president of Clark University, invited both Freud and Jung to come to the United States to lecture on the occasion of the twentieth anniversary of Clark University. Both by attendance and by the subsequent publication of these lectures in the *American Journal of Psychology* (51) psychologists became more familiar with their work. In the meantime, Brill (23) was translating Freud's works, and other psychoanalysts began practice. By 1911 there was enough interest that the first psychoanalytic association, the New York Psychoanalytic Society, was founded.

In view of these factors in the history of clinical psychology, it is possible to offer the interpretation that actually it was partly the psychologists and not psychiatrists alone, as is commonly supposed, who offered the first support to psychoanalysis in the United States. To be sure, in the twenties the psychiatrists in increasing numbers became interested and during the following twenty years became so firmly identified with the field that it is only today that psychologists, as psychologists, are again beginning to assume any prominence in psychoanalytic thinking and practice.

The Psychologist and the Psychological Clinic

It has been accepted by psychologists quite generally that the case leading to the founding of the first psychological clinic was treated by Lightner Witmer (114) at the University of Pennsylvania in March 1896.

[2] Before taking a medical degree, Boris Sidis, then a psychologist, published in 1907 in the *Boston Medical and Surgical Journal* (96) a series of cases of what he called hypnoidal states treated by his particular method of suggestion.

Witmer was the first to speak of the "psychological clinic," of "clinical psychology," and the "clinical method in psychology" (26). The history of his clinic has been discussed elsewhere (26, 27, 93, 107, 114) and is quite well known. It is, therefore, unnecessary to dwell upon it. Instead, after very briefly examining its functioning, attention will be given to the extent of its influence upon the history of clinical psychology.

Even a cursory examination of the early issues of the *Psychological Clinic*, a journal founded and edited by Witmer, will show that the work attempted in this clinic included referral to medical sources, the presence of social workers, and many other "modern" innovations discussed by the writer elsewhere (107). On the other hand, although the juvenile court and social agencies referred cases to Witmer's clinic, the great majority came from the school system. Much attention was paid to the relation of physical defects and neurological conditions to behavior problems. Cooperation with special teachers of the blind and deaf and the mentally defective was stressed. In general, intellectual aspects of children's problems were emphasized, using a biographical approach. Relatively few psychologists published in the *Psychological Clinic* in the early years. Educators, either teachers, principals, or professors, wrote the majority of the articles during this period. In later years the publications of psychologists predominated. The articles are chiefly of antiquarian interest today.

The clinic founded by Seashore (91) at the University of Iowa about 1910 was modeled after Witmer's clinic, and others, such as that founded by J. E. W. Wallin of the University of Pittsburgh in 1912, undoubtedly owe part of their impetus to it, but many other psychological clinics and activities seemed to grow up independently and with little knowledge of the development of this first clinic (97). For example, Seashore (91) speaks of his as the "second" psychological clinic. And yet in 1914 Wallin (105) found about 20 psychological clinics to be in existence, of which some at least must have developed under a different tradition except in the rather unlikely event that the great majority were founded after 1910, but before 1914. Although the Witmer clinic has been functioning continuously since its inception, it is quite difficult to find evidence of its effects upon clinical psychology today. This has not been due to lack of local support; rather it is because its influence did not spread beyond Philadelphia to any considerable degree. The reasons for this relative lack of influence will be discussed after considering a related development: the child guidance movement.

The Psychologist in Child Guidance

Still another stream which merged into the torrent that is clinical psychology today came from the so-called child guidance movement. In this effort William Healy (59), a psychiatrist, was the most important early figure. The beginnings of this movement arose from the conviction that antisocial behavior was treatable by psychiatric means. A subsequently discarded tenet which went hand in hand with this conviction was an emphasis upon pathology. Hence the first "child guidance" clinic, at the time of its founding in Chicago in 1909, was called "The Juvenile Psychopathic Institute." It is perhaps prophetic that the selec-

tion of Healy for the position of director was "as a pupil of James and a free lance in competition with a more rigid Wundtian and experimentally and statistically minded psychologist" (76, p. 242). Its first staff was very small, consisting of Dr. Healy, as psychiatrist, Dr. Grace M. Fernald, as psychologist, and one secretary. It is important to note that no social worker was a paid member of the staff, but Healy indicates that social workers from cooperating agencies worked with them from the very beginning. Only later did the specialty of psychiatric social worker, as such, emerge. Mental testing by Fernald, and later by Augusta F. Bronner, emphasized performance testing and other instruments of local origin. In 1910, however, Healy introduced the Binet-Simon tests into the United States (as did Goddard at Vineland simultaneously and independently). A direct outgrowth of the use of this and other instruments was the publication in 1927 of a *Manual of Individual Tests and Testing* (25) by Bronner, Healy, and their co-workers. Both Healy and Bronner had migrated eastward, organizing in 1917 a clinic in Boston under the name of the Judge Baker Foundation, later changed to the Judge Baker Guidance Center. This venture was enormously successful and resulted in still further important work in the field of delinquency. Many publications, including several books upon problems of the delinquent, had considerable influence upon patterns in this field.

In contrasting the relative success of Healy's venture and its continuity with the present with the relative lack of influence of Witmer's clinic, Shakow (92) presents a thoughtfully detailed statement, one or two points of which can be mentioned. The psy-

chologist Witmer was concerned with intellectual aspects of the functioning individual, worked primarily with mental defectives or school retardation problems, when concerned with medical aspects focused more on the physical or neurological, and, most important of all, identified himself with the Wundt-Kraepelin point of view. On the other hand, the psychiatrist Healy was concerned with affective aspects of the personality, worked primarily with behavior problems and delinquency, when concerned with medical problems stressed the psychiatric, and, again most important of all, was profoundly influenced by James and Freud. Although a pioneer, Witmer turned his back on almost all that was to predominate in the later days of clinical psychology and became of historical significance only. Healy is still a contemporary.

The Psychologist in Mental Hospitals

The importance of McLean Hospital in Waverly, Massachusetts has never been fully appreciated in the history of psychiatry and psychology. Founded in 1818, its superintendent at the turn of the century was Dr. Edward Cowles, a former surgeon in the Union Army.[3] Years later he took some incidental training in psychology at Johns Hopkins (57). In many ways he was a man ahead of his time. He encouraged research and brought to this hospital biochemists, pathologists, physiologists, and psychologists. One could date the beginnings of conjoint medicine as taking place at McLean Hospital since these approaches were used in its laboratory sometime before 1894. In that year Hall described the laboratory as fol-

[3] SHAFFER, P. A. Personal communication, 1952.

lows: "The work of this laboratory was begun in 1889, for the clinical purposes of the hospital. It is sought to combine neurological studies in the departments of psychiatry and physiological psychology, and their relations with anatomical and chemical pathology, etc." (57, p. 358). Only a quotation from Cowles will bring out the contemporary ring of his words:

The purpose of establishing and developing the laboratory has been carried on under much difficulty, naturally due to the newness of the attempt to combine with psychiatry the other departments of scientific medical research. The pathology of the terminal stages of insanity must be studied as heretofore, and it is necessary to add that of the initial conditions which lead to mental disorder. Such studies must therefore be combined with physiological psychology in the attempt to determine the exact nature and causes of departures from normal mental function. Also, in the dependence of these changes upon general physiological processes, and in order to take into account all the elements of vital activity involved, it is supremely necessary to study both physiological and pathological chemistry in their direct and indirect relations to mental changes (57, p. 363).

Research efforts along these lines apparently first emerged from this laboratory. In presenting the history of psychiatric research, Whitehorn (111) recognized this contribution and first described McLean Hospital and its work before dealing with any other developments.

Cowles, in a review of the progress in psychiatry at the time of the fiftieth anniversary of the American Psychiatric Association in 1894, emphasized the importance of what he referred to as the systems of "new psychology" as one of the "most hopeful signs of progress" to bring about advancement in the under-

standing of mental diseases (34). Either as frequent visitors from nearby Boston or as members of the staff of McLean Hospital at this time were Morton Prince, August Hoch, Boris Sidis, and Adolf Meyer. Interest in psychology is shown by the fact that Cowles and William Noyes, of the same hospital, were among the approximately 13 to 18 individuals who were present at the founding of the American Psychological Association at Clark University in 1892 (36).

In 1893 August Hoch (75) was selected by Cowles to be psychologist and pathologist at McLean. The use of the term *psychologist* was neither idle nor esoteric. Having previously received a medical education, he now was sent abroad for further training, and it would appear that much of his training was in psychology with Wundt, Kulpe, Marbe, and Kiesow. He also worked with Kraepelin. On assuming his post at McLean he turned to work with the ergograph in clinical problems and in the first volume of the *Psychological Bulletin* (62) summed up experimentation in this field. Subsequently, as professor of psychiatry at Cornell and director of the Psychiatric Institute, he turned to more narrowly psychiatric problems, but there would appear to be little doubt that during this period at McLean he functioned, in part at least, as a psychologist.

It was in this atmosphere that a psychological laboratory was founded. This laboratory was begun in 1904 at McLean Hospital by Shepard Ivory Franz (50). It was influential in the *rapprochement* of psychology and psychopathology, although often interested in matters more physiological than psychological. The laboratory became estab-

lished under the direction of Franz, and on his leaving for what is now St. Elizabeths Hospital of Washington, F. Lyman Wells was appointed his successor and remained there until 1921.

Franz continued his interest while in Washington, not only writing such articles with a modern ring, although published in 1912, as "The Present Status of Psychology in Medical Education and Practice" (25), but also introducing in 1907 a routine clinical psychological examination of all new patients in a mental hospital setting. This was probably the first instance of routine psychological testing of psychiatric hospital patients. Among Franz's associates during the early period were Grace H. Kent and Edwin G. Boring, both of whom published on learning in dementia praecox. Although Boring, as is well known, returned to other fields, he nevertheless felt that the summer he spent in the hospital was a very valuable, broadening experience (22). From 1906 to 1921 Grace H. Kent was psychologist at Philadelphia Hospital, Kings' Park State Hospital, and St. Elizabeths, respectively. In 1922 she went to Worcester State Hospital, remaining there until 1926 (79). Thereafter, for many years she was at Danvers State Hospital.

The Psychologist and Institutions for the Mentally Defective

It was Goddard's laboratory at the Vineland Training School that was the second center to be devoted to the psychological study of the feebleminded.[4] Henry H. Goddard became director of psychological research at this institution in 1906 and was influential in the establishment of the

[4] In 1898 Wylie, a physician, began psychological testing at the state institution for the feebleminded at Faribault, Minnesota (98).

psychologist as a person working with the mentally defective. As mentioned earlier, he first translated and used the Binet in this country. For practical purposes, the use of the Binet was at this time almost exclusively restricted to the feebleminded. It was from this center that the Binet spread to other institutions (84). His directorship continued until almost the twenties.

Psychology as a Profession

It was as early as 1904 that Cattell (30) made the prediction that there would eventually be a profession as well as a science of psychology. Actually professional action preceded this pronouncement.

For purposes of this presentation the relevant characteristics of a profession include establishment of commonly agreed-upon practices concerning relationship with colleagues and with the public served. The questions of competency and the means of controlling competency immediately arise. Traditionally, a profession controls competency among its own members. Thus, self-determined control of its members is the hallmark of a profession.

The first stirrings of attempts at control arose in the American Psychological Association and took the form of considering control of clinical procedure through evaluation of test data. In 1895, only three years after the founding of the Association, J. Mark Baldwin, in the words of Fernberger, "proposed the formation of a committee to consider the feasibility of cooperation among the psychological laboratories for the collection of mental and physical statistics" (43, p. 42). The committee that was appointed, chaired by Cattell, called itself "The Committee on Physical and Mental Tests," but the battery

of tests they proposed for try-out to develop norms gained little acceptance so that after 1899 no further word was heard from this committee. Another committee for the purpose of establishing methods of testing was appointed in 1907 and continued until 1919. It made some progress, for example, sponsoring research on the Woodworth-Wells Association Tests, but it fell far short of the ostensible goal.

In 1915, on the motion of Guy M. Whipple, the Association went on record as "discouraging" the use of mental tests by unqualified individuals. In 1917 a committee to consider qualifications for psychological examiners was appointed, and two years later one to consider certifying "consulting" psychologists. In 1919 the Section of Clinical Psychology within the American Psychological Association was formed (43). In large measure, it was a "special interest group" concerned with arranging programs at the annual meetings and the like. Its members were, however, drawn into the discussion, pro and con, of the merits of certification. After much maneuvering, favorable action on certification of clinical psychologists finally resulted, and the first certificates were granted after the 1921 meeting. However, only twenty-five psychologists applied, and the project was abandoned. The death blow was dealt by an APA policy committee which considered that certification was not practicable and, on vote of the APA membership in 1927, discontinued certification. In some measure at least, the decision was influenced by the realization that with certification went the problem of enforcement of the standards instituted, especially on psychological workers outside the membership. Thereafter, according to Fernberger (43), there was a period of some years

without important action within the American Psychological Association on these problems.

Internship training, as distinguished from academic course work, is a manifestation of professional training. Morrow (78) indicates that Lightner Witmer was apparently the first to suggest practical work for the psychologist through training school and laboratory. However, the first actual internships were those offered by the Training School in Vineland, New Jersey, under the supervision of H. H. Goddard. This program began in 1908 and has continued down to the present time. In 1909 William Healy began accepting graduate students at the Juvenile Psychopathic Institute in Chicago, while the first internship in a psychiatric institution for adults was established in 1913 at the Boston Psychopathic Hospital under the direction of Robert M. Yerkes. Other earlier internships include those at Worcester State Hospital, McLean Hospital, the Western State Penitentiary in Pennsylvania, and the New York Institute for Child Guidance.

CLINICAL PSYCHOLOGY IN THE TWENTIES AND THIRTIES

In the twenties and thirties clinical psychology left the period of its lusty, disorganized infancy and entered its rather undernourished but rapid and stormy adolescence. As late as 1918 only 15 members or 4 per cent of the APA listed the field of clinical psychology as a research interest (44). This rose to 99 members or 19 per cent in 1937. In that year the newly instituted membership category of Associate showed 428 or 28 per cent interested in clinical psychology, the largest field of interest for this class of membership. In increasing numbers clinical psychologists were employed in hospitals,

clinics, schools, penal institutions, social agencies, homes for the feeble-minded, industrial plants, and the entire gamut of agencies concerned with human welfare. For example, Finch and Odoroff (46), in a survey concerning employment trends, indicate that of 1,267 members of the American Psychological Association in 1931, 286 or 26.9 per cent were not in teaching positions. In 1940 the number of nonteachers had swelled to 888 or 39.3 per cent of the membership. Clinical nonteachers increased from 95 to 272 during this ten-year period.

It was during this period that many psychologists did yeoman service for clinical psychology without being primarily identified with the field. Carl E. Seashore (91) may be used as an illustration. It has already been noted that he founded a psychological clinic at the University of Iowa about 1910. During the period now under consideration he was interested in the relationship between psychology and psychiatry and took the lead in organizing a national conference on this topic. He also aided in founding the Iowa Psychopathic Hospital and worked with Samuel Orton, Edward Lee Travis, and Wendell Johnson in speech pathology. Many other men such as Gardner Murphy, Goodwin Watson, Horace B. English, Albert T. Poffenberger, Kurt Lewin, Carney Landis, Robert M. Yerkes, Walter R. Miles, Gordon W. Allport, and Kurt Goldstein, although primarily associated with some other aspect of psychology, also performed services for the clinical field.

In spite of such developments as those just described, Loutitt (70) could indicate during the same period that "American Psychology, generally speaking, has not been greatly interested in practical problems of human behavior" (70, p. 361). This contention applied to clinical psychology with as much force as, or more than, it did to other applications of psychology. Most of the difficulties that clinical psychology went through during this period as it groped toward professional stature were internal to the field itself. Both rapid growth and some hostility from the dominant entrenched forces in psychology are imbedded in the history of the period and influence many of the specific developments now to be discussed.

The Psychometric Tradition

The period of the twenties was, in the words of Merrill, a "plateau . . . [following] the initial impetus given to testing when these first tools of the clinician were being subjected to evaluation, and the exaggerated expectations of over-enthusiastic users were being reduced in the crucibles of research" (73, p. 283). Studies of validity, investigations of the constancy of IQ, application of the tests to new populations, studies of individual differences, the nature-nurture controversy, racial differences, the development of group testing, achievement tests, interest measures, and personality testing of the questionnaire variety occupied this and the following decade and helped to consolidate the gains of the previous period. Theories of intelligence and factor analysis are also intimately related to this trend. It was a period, as the term plateau implies, of masked gain which prepared the way for the present period.

More and more objections began to be raised to the limitations entailed by this approach. The development of group tests during and after World War I placed a premium on easy reproduction, rigid standardization down to the slightest detail,

and emphasis on the score obtained to the exclusion of all else. Measures of personality with these same characteristics were developed during the twenties and thirties. To some psychologists the results obtained were considered disappointing and sterile.

In 1927 F. L. Wells published *Mental Tests in Clinical Practice* (108), in which he stated vividly the major objection to a rigid psychometric approach:

An intelligent South Sea Islander, observing a psychometric examination, would be likely to regard it as a magic rite designed to propitiate friendly spirits in the patient's behalf. Should he observe a conscientious examiner in the apprentice stage, tightly clinging to forms prescribed, his idea would be confirmed, for none knows better than himself how slight a departure from the required formulae will not only destroy their beneficence but may well deliver the hapless sufferer into the hands of the malignant ghosts. Over against such esoteric views of psychometric methods is the customary and pragmatic one. The function of psychometrics is not the accomplishment of a ritual, but the understanding of the patient. The ceremony of mental tests is valuable so far as it serves to reach this end. When it fails, or stands in the way of doing this, proper technique demands that it be modified. Ability to do this intelligently is what distinguishes the psychologist, properly so called, from the "mental tester" (108, p. 27).

Further objections to exclusive reliance upon a psychometric approach arose with the emergence of projective techniques as an aspect of the dynamic tradition next to be considered.

The Dynamic Tradition

Many of the present developments in clinical psychology—the emphasis on understanding of personality functioning, the attempt to relate present behavior to experiences of which the patient is unaware, the evaluative use of incidental verbalizations and physical behavior of the patient, and the artistic element in psychodiagnostic appraisal—stem in large measure from the dynamic tradition.

In terms of the sources of these influences, Sigmund Freud, of course, looms largest. He and his fellow analysts profoundly affected the thinking of many clinical psychologists, who were for the most part passive recipients of this influence. No longer did they share leadership with their medical colleagues as during the first twenty years of the century. The influence of psychoanalysis was felt directly on three of the specific manifestations of the dynamic tradition directly involving the psychologist—projective techniques, the Harvard Psychological Clinic, and the American Orthopsychiatric Association. The first, an approach to personality, the second, a clinic, and the third, a professional organization, share responsibility as the most important manifestations of the dynamic tradition in psychology of the day. Each will be considered in turn.

Hermann Rorschach, a Swiss psychiatrist, published with Oberholzer on the specific but intricate relationships which exist between his inkblot technique and psychoanalysis. The technique itself occupied much of his time between 1911 and his untimely death in 1922. In the United States pioneering work with the Rorschach was done by David M. Levy, a child psychiatrist, with whom Samuel J. Beck became associated beginning in 1927. In 1930 Beck presented the first Rorschach study in this country as his doctoral dissertation, and during the thirties the Rorschach technique came more and more into

prominence in clinical circles. Along with Beck, pioneer American psychologists who made major contributions to Rorschach literature during this period were Bruno Klopfer, Marguerite Hertz, and Zygmunt Piotrowski.

Two reasons are given by Beck (18) for the increasing preoccupation of psychologists with the Rorschach technique as compared to psychiatrists. There is, first, the division of labor with the Rorschach as one of the diagnostic testing instruments and, second, the fact that its use spread outside the narrowly psychiatric area into the schools, work with delinquents, industry, and the like. Other projective techniques, notably the Thematic Apperception Test, also appeared during the thirties.

In the meantime the psychodynamic emphasis began to be a part of the intellectual armamentarium of the psychologist. The article by L. K. Frank, "Projective Methods for the Study of Personality" (48), published in 1939, offered a rationale for the projective approach and stimulated both research and theoretical efforts in the decades to follow. Merrill summarizes other reasons for the rapid spread of projective testing as having

. . . had to do with significant changes that have been occurring in the clinician's self-concept and his changing perception of his role as his social responsibilities grow and expand. Projective tests have become important tools for the psychotherapists. These tests command the attention and respect of our colleagues in the medical fraternity, the psychiatrists. They constitute moreover, the basic technological structure upon which is being built a new systematic point of view, projective psychology with its own theory of personality. This new projective psychology has been aptly characterized as a psychology of protest. As both be-

haviorism and Gestalt psychology came about as protests against the established psychologies called structural, so this emerging projective psychology runs sharply counter to the traditions that have characterized individual psychology in America. Having something to push against, it can move (73, p. 286).

In 1927 the Harvard Psychological Clinic was founded by Morton Prince. Its express purpose was to bring together academic and clinical psychology. Henry A. Murray took over headship of the clinic early in the thirties and with a large group of collaborators, including Donald W. MacKinnon, Saul Rosenzweig, R. Nevitt Sanford, and Robert W. White, carried on a brilliant research project in personality functioning. This culminated in 1938 in the well-known *Explorations in Personality* (81).

The American Orthopsychiatric Association is an organization with ties to child guidance in particular and to the dynamic tradition in general. It was founded in 1924 with many of the leaders of the child guidance movement present (71). William Healy was elected president during this year and served through 1926. Later presidents included Karl Menninger, David Levy, and in 1931, the first psychologist to be president, Augusta F. Bronner. Other psychologist presidents were Edgar A. Doll, Samuel Beck, and Morris Krugman. After thinking through the problem of membership, originally restricted to psychiatrists, the pattern emerged in 1926 of having as full members "psychiatrists, psychologists, and other professional persons whose work and interests lie in the study and treatment of conduct disorders" (71, p. 199). Both the letter and spirit of this method of organization for work interchange, support, and

advance have continued to the present day. However, there has never been any question, as might have been foretold from the original organization, but that psychiatrists were dominant in it. For example, twenty-one of the first twenty-six presidents held the M.D. degree, only four being psychologists and only one a social worker. This organization continues to wield much influence both through its journal, *The American Journal of Orthopsychiatry*, and through its annual meetings which are characteristically attended by far more nonmembers than members.

Psychological Clinics

This was a period during which psychological clinics reflected the plateau of the psychometric tradition. Some new clinics appeared; others closed their doors (70). In 1934 a survey report of a questionnaire of psychoeducational clinics by Witty and Theman (115) appeared. On the basis of their returns they estimated that there were about 50 such clinics. This figure may be contrasted with the approximately 20 found by Wallin in 1914 (105). In 1932 the median length of time the clinics had been in existence was four years. Located in colleges, universities, teachers colleges, and normal schools, their stated purposes involved (*a*) providing schools, social agencies, and individuals with diagnostic test services and remedial methods in order to bring about educational, vocational, and social adjustment; (*b*) training students in giving and interpreting tests; and (*c*) research with emphasis on the study of deviates, causes and treatment of learning difficulties, and work with remedial materials. It would appear that this survey epitomizes the work of psychological clinics of the day, featuring emphasis on testing and remedial education.

Child Guidance

The period 1922–1927 was one in which the National Committee for Mental Hygiene on behalf of the Commonwealth Fund established demonstration clinics in a variety of cities and rural areas for the purpose of showing both their need and the work they could do (99). For the first time they were called "child guidance clinics." Eight clinics were permanently established directly as a result, and many others were at least partially stimulated by this effort. It was the announced intention from the very beginning that eventually expenses for their maintenance would be absorbed by the community in which they were located. Deliberate experimentation as to method of organization in relation to other agencies was carried out—some were attached to the courts, others to local charities, to university and to teaching hospitals. The child guidance clinic plan of organization called for the professional personnel to include at least a psychiatrist, a psychologist, and a social worker. These activities in their formative stages continued roughly over the decade 1920–1930. To be sure, their influence and organization continued thereafter, but this period marked the heyday of their unique contribution.

An important shift of focus of attention had been occurring during this period. No longer was the delinquent of primary interest. Nor was there much concern with mental defectives, epileptics, or neurological cases. Instead, emphasis was placed upon maladjustment in school and home, especially that centering around parent-child relationships.

The clinics began to concentrate upon problems of the individual who may be spoken of as falling within the normal range of intelligence, the roots of which may in some measure be traced to emotional difficulties.

The Psychologist in Mental Hospitals

In 1921 Wells left McLean Hospital for Boston Psychopathic Hospital where he served as head psychologist until 1938. This hospital also became a center of clinical activity and training. A pioneer in present-day clinical psychology, David Shakow, now of the Illinois Neuropsychiatric Institute, is still very active. For a period of nearly twenty years, Shakow served as director of psychological research at Worcester State Hospital. His activities, along with his research efforts, included direction of an internship training program. It apparently was the closest in spirit to the modern internship program, and his experience derived in this setting was of great value in formulating present-day practices concerning internship.

A gradual increase in the number of clinical psychologists in mental hospitals was taking place. However, the geographical isolation of most such hospital psychologists apparently accentuated an isolation on other grounds so that the effect of this aspect of the development of clinical psychology was not as important as it was to be in the decades to come. Nevertheless, some psychologists were beginning to suspect that their approach was unduly limited. As a result, there were serious attempts at broadening the scope of testing efforts, to escape the atomistic tradition by means of research and theorizing concerning the personality of their patients.

Mental Deficiency

In 1919 Goddard was succeeded by Stanley D. Porteus as director of the Vineland Laboratory. Under both his direction and the subsequent direction from 1925 until 1949 of Edgar A. Doll the clinical problems of feeblemindedness received intensive and extensive study.

It was precisely in the field of intelligence testing that clinical psychology was most advanced during this period. Psychometric testing of suspected mental deficiency was widely accepted, and the psychologist was the authority in this field (28). Nevertheless, as Buck (28) indicates, appreciation of the complexity, rather than the simplicity, of the diagnosis of mental deficiency emerged during these two decades.

The fact that a person was doing clinical psychological work with mental defectives unfortunately indicated almost nothing about the nature of the training and experience of the practitioner in question. In 1940 Hackbusch (56) reported an inquiry concerning psychological work in state and private institutions for the mentally defective. Of the approximately 100 institutions which replied apparently all were doing some form of psychological testing. However, less than half had a psychologist on their staff. The remainder had their testing done by outside sources, or by teachers, social workers, and physicians on their own staffs. It is also noteworthy that the "psychologists" apparently varied widely in the nature of their background. Some had less than an A.B. degree, while others had an A.B. or an M.A. in addition, but very few held the Ph.D. degree. Therefore, despite the acceptance of their work, the status of psychologists and psychological work in the twen-

ties and thirties was somewhat confused.

Psychology as a Profession

The origins of professional activity, as has been indicated, were centered within the American Psychological Association. This period extended from 1895 to the mid-twenties. Founded to advance psychology as a *science*, the Association had not been singularly successful in reflecting the interests of its members either in applications of psychology or in their professional aspirations. The twenties and thirties were characterized by the advent of other organizations more directly concerned with professional problems.

In 1917 a group of psychologists interested in the advancement of the practice of psychology met in Pittsburgh, Pennsylvania. Leta S. Hollingworth took the initiative in bringing the group together, and prominent charter members included Bronner, Fernald, Healy, Kuhlmann, Pintner, Terman, Whipple, Wells, and Yerkes. To quote Symonds, "After a brief history of two years, during which a bitter struggle went on in the American Psychological Association over the question of authority for certification of psychologists for clinical work, the American Association of Clinical Psychologists became defunct through the adoption by the APA of a report recommending the establishment of the AACP as a Section of Clinical Psychology" (100, p. 337). According to the same writer, the next step was the slow development of various local groups concerned with applied and professional matters in several states.

In 1930 the Association of Consulting Psychologists was reorganized from a still earlier association founded in 1921 (40, 52). Gradually it extended its membership beyond New York and environs and became one of the more important elements later to merge into the American Association for Applied Psychology (AAAP). The organization meeting of this association took place in 1937. Many of the difficulties in organizing centered upon the standards for membership. Then, as now, there was the dilemma of maintaining standards and yet not setting them so high as to exclude the majority of those doing work in the applied fields. Eventually this was settled, and a national organization concerned with all aspects of the application of psychology came into being and became the dominant national professional organization. A divisional structure was followed with clinical, educational, industrial, and consulting sections.

The Journal of Consulting Psychology was at first a publication of the Association of Consulting Psychologists and then of the AAAP. Papers in clinical, educational, industrial, and consulting psychology appeared, but a considerable portion of space was devoted to organizational and professional matters (100).

Thus, there existed at the close of the thirties two major psychological societies—one dedicated to the advancement of psychology as a science and the other to its application. Generally speaking, members of the latter also had membership in the former but sincerely felt the essential nature of their applied organization. The Psychometric Society and the Society for the Psychological Study of Social Issues were also founded during this period. In part at least, these organizations arose because of similar dissatisfaction with the adequacy of representation of some of

their interests in the American Psychological Association. So the thirties closed with at least the possibility of dangerous rifts in the ranks of psychologists. However, as is well known, this danger passed in the forties with all of these organizations integrated into the reorganized American Psychological Association (116). In 1945 this reorganization went into effect. Both in spirit and in practice the American Psychological Association represents psychology as a science and as a profession.

The Psychologist and Therapy

During the twenties and thirties there appeared to be a gradual increase in the number of clinical psychologists engaging in therapy. From the time of Sidis and Scott at the turn of the century some psychologists had been so employed. In many instances psychotherapeutic practice grew out of the psychologist's educational function. Considered as expert both in matters of learning as a subject of investigation and in education as a field of endeavor, the psychologist worked with patients, particularly children, with whom remedial education was necessary. A similar process took place to a lesser extent in psychiatric clinics. It was in the hospitals that this development lagged, partly because the sheer press of numbers of patients confined the psychologists to psychodiagnostic tasks and partly because psychotherapy, except at a few institutions, was not practiced at all. There was relatively little difficulty in interprofessional relations with psychiatry during this period. In large measure this was because there were few psychologists practicing therapy, and these few were doing so under institutional auspices and exceptional circumstances. Then too,

the psychiatrist himself was more isolated both from his medical colleagues and from the public than he is today. More concerned with the psychotic and the adult than with the neurotic and the child, his path did not as often cross that of the psychologist as it did in the forties and fifties.

No continuity in the development of psychotherapists among psychologists is discernible from generation to generation. Neither Sidis nor Scott stimulated psychologists to work with psychotherapeutic problems. In later years individual psychologists prominent in psychotherapy gained in stature, not unaided to be sure, but also not from the combined efforts of any group or from the work of one senior individual. Phyllis Blanchard, for example, an acknowledged leading therapist, as attested to by her presence in leading symposia and by books on the topic, neither received her training in therapy from psychologists nor participated in the training of psychologists in therapy. Other therapist-psychologists, also, developed along individual lines. The work of Carl Rogers, although begun in the thirties, did not reach national prominence until the forties.

PSYCHOLOGISTS IN THE ARMED SERVICES AND THEIR EFFECT UPON PSYCHOLOGY IN THE POSTWAR PERIOD[5]

About 1,500 psychologists served in the armed services during World War II. About one out of four psychologists thus was called upon to function in an applied field—that is, psychology applied to the very

practical problem of war. Moreover, this group was predominantly young, averaging about 32 years of age (24), thus including many individuals just reaching professional maturity. It is not unduly optimistic to suppose that some of their experiences during these tours of duty carried over in attitude and practice to the postwar years.

To appreciate properly certain changes of attitude, it must be remembered that a considerable number of psychologists in uniform were products of an academic tradition whose isolationist tendencies in regard to professional application prior to the war they were quite willingly and even complacently furthering. In fact, Andrews and Dreese (16) found that almost 90 per cent of the psychologists in military service were in academic or governmental work prior to the war.

From the process of learning to apply their psychological training to the military situation, later consideration revealed at least two major trends that have had, and will continue to have, profound effect upon contemporary psychology. They discovered to their mild surprise, and to the considerable amazement of their colleagues from other disciplines, that their general training in psychological methods was capable of application to many problems which at first seemed utterly alien to their background. From aircraft instrument-panel design to selecting underwater demolition teams, psychologists found that they, in collaboration with specialists from other fields, had something valuable to contribute. Realization was forced upon them that an experimental background in psychology is capable of transfer to intelligent and capable handling of many sorts of problems.

Paradoxically, however, they gained added respect for the clinical approach. In this connection it must be realized that almost half of the psychologists used clinical and counseling procedures during some part of their period in uniform (24). Many psychologists, willy-nilly, were placed in a position where they functioned in selection and assignment, sat as members of discharge boards, worked as members of clinical teams, conducted therapeutic sessions, both group and individual, and in these and many other ways used diagnostic and treatment methods. Concrete expressions of this interest can be found in an article by Britt and Morgan (24) concerning the results obtained from a questionnaire mailed to every psychologist in uniform. They conclude that there was an overwhelming interest in having more practical postwar graduate training. Nearly 24 per cent of the suggestions for new courses for graduate study were clearly within the general clinical field. At least some of the armed services psychologists who had previously not been particularly receptive came to understand and appreciate the contributions, past and potential, of the clinical method. This impression is verified by the finding in a survey by Andrews and Dreese (16) that three times as many military psychologists engaged in clinical work after the war as had done so in the prewar period.

CLINICAL PSYCHOLOGY TODAY

With the coming of the forties and World War II, one leaves the realm of history and enters the present. It would be both hazardous and presumptuous to attempt to trace in detail the events from this time on. Nevertheless, certain factors in the foregoing account may be related to

present trends. Clinical psychology as a method, as an attitude, and as a field of endeavor is reflected in its past.

It would appear that clinical psychology and academic psychology have influenced each other markedly, with a reciprocal, symbiotic relationship having been formed. Other disciplines, notably the medical and particularly the psychiatric and psychoanalytic, influenced and vitalized clinical psychology.

Through the thirties certain predominant aspects may be referred to as "child," "psychological," and "clinical" as contrasted with "adult," "psychiatric," and "institutional" functions. Distinctively psychological clinics and work with children are not only important because of their service and scientific value, but also for the community orientation that they manifest and the preventive emphasis that they maintain. And yet since the thirties the emphasis has shifted.

The "adult," "psychiatric," and "institutional" aspects of clinical psychology appear to be dominant today, but this is by no means an unmixed blessing. Many of the more vocal leaders of the field, including to some extent the official committees of the American Psychological Association, have fostered emphasis upon the former. The extremely valuable support rendered by the Veterans Administration to our training and practice has emphasized the current trend. Work with adults in a psychiatrically oriented institution is a specialty, albeit an important one, in the broader field.

With the forties also came the domination in the history of clinical psychology of one of the trends previously sketched. This was the emergence and implementation of a

concept of a profession of psychology. One illustration will suffice. Until after World War II, there was relatively little demonstrable agreement about the training, nature, duties, or status of the clinical psychologist. To quote Eysenck, "A person who called himself a 'clinical psychologist' might be someone of great eminence, highly qualified academically and with 20 or 30 years of practical experience in the fields of diagnostic testing, research, and therapy, or he might be a student just graduated from the University, without any kind of relevant experience, capable only of grinding out Binet I.Q.'s without even an adequate understanding of their relevance to the clinical problem presented" (41, p. 711). The facet of the professionalization of a psychologist, although not completely defined today, has reached a degree of precise formulation undreamed of a few years ago.

Current issues and accomplishments, stabilizing trends, and unresolved problems may be related to the emergence of psychology as a profession. Factors making for the present stabilization include the agreement of the great majority of interested parties concerning diagnostic appraisal as a task of the clinical psychologist (7, 9, 12, 55, 94), the present organization and function of the American Psychological Association (4, 116), current efforts directed toward the training of clinical psychologists (9, 10, 12, 14), present activities looking toward codification of ethical problems (5, 6, 17, 21, 61), and the influence of such institutions as the American Board of Examiners in Professional Psychology (2, 3), state societies (11), the United States Public Health Service (42, 104), the Veterans Administration (1, 58, 103),

and the armed services (102). On the other hand, currently unresolved issues face the profession today. The problems on which there are differences of opinion both in psychology and in other professions include psychotherapy as a task of the psychologist and the nature of the relation of psychology to psychiatry and medicine (7, 13, 54, 55, 72, 87, 94), the nature of the relation of psychology to social work (33), the question of the advisability of certification and licensure (31, 47, 60, 109, 112, 117), the question of the desirability of private practice (38, 39), the position and function of non-Ph.D.'s in clinical psychology (15, 20, 35, 67, 69, 98), and the "imbalance" in psychology between scientific and professional demands (63, 87, 88, 89). Not only do these problems have roots in the past, but they are also an expression of the period of professionalization of large segments of psychology today.

World War II focused the needs and demonstrated what could be done in clinical psychology; the period after the war is still feeling the pressure of these social needs and is witnessing the reactions, adaptive and otherwise, of a beginning profession to these demands.

REFERENCES

1. ADLER, M. H., FUTTERMAN, S., & WEBB, R. Activities of the mental hygiene clinics of the Veterans Administration. *J. Clin. Psychopath.*, 1948, 9, 517–527.
2. AMERICAN BOARD OF EXAMINERS IN PROFESSIONAL PSYCHOLOGY, INC. *Official Bulletin.* 1948, No. 1.
3. AMERICAN BOARD OF EXAMINERS IN PROFESSIONAL PSYCHOLOGY. The work of the American Board of Examiners in Professional Psychology: annual report of the Board to the members of the APA. *Amer. Psychologist,* 1951, 6, 620–625.
4. AMERICAN PSYCHOLOGICAL ASSOCIATION. By-laws for the American Psychological Association (as amended through September, 1951). In *Directory, American Psychologicl Association.* Washington, D. C.: American Psychological Association, 1951.
5. AMERICAN PSYCHOLOGICAL ASSOCIATION. *Ethical standards for psychologists.* Vol. 1: *The code of ethics.* Washington, D. C.: American Psychological Association, 1952.
6. AMERICAN PSYCHOLOGICAL ASSOCIATION. *Ethical standards for psychologists.* Vol. 2: *Source book of ethical problems, incidents, and principles.* Washington, D. C.: American Psychological Association, 1952.
7. AMERICAN PSYCHOLOGICAL ASSOCIATION, *Ad Hoc* COMMITTEE ON RELATIONS BETWEEN PSYCHOLOGY AND THE MEDICAL PROFESSION. Psychology and its relationships with other professions. *Amer. Psychologist,* 1952, 7, 145–152.
8. AMERICAN PSYCHOLOGICAL ASSOCIATION, COMMITTEE OF CLINICAL SECTION. I. The definition of clinical psychology and standards of training for clinical psychologists. II. Guide to psychological clinics in the United States. *Psychol. Clin.*, 1935, 23, 2–140.
9. AMERICAN PSYCHOLOGICAL ASSOCIATION, COMMITTEE ON TRAINING IN CLINICAL PSYCHOLOGY. Recommended graduate training program in clinical psychology. *Amer. Psychologist,* 1947, 2, 539–558.
10. AMERICAN PSYCHOLOGICAL ASSOCIATION, COMMITTEE ON TRAINING IN CLINICAL PSYCHOLOGY. Annual report of the Committee on Training in Clinical Psychology. *Amer. Psychologist,* 1951, 6, 612–617.
11. AMERICAN PSYCHOLOGICAL ASSOCIATION, CONFERENCE OF STATE PSYCHOLOGICAL ASSOCIATIONS. *CSPA Newsletter,* April, 1952. (Mimeo.)
12. AMERICAN PSYCHOLOGICAL ASSOCIATION, CONFERENCE ON GRADUATE EDUCATION IN CLINICAL PSYCHOLOGY, Boulder, Colorado. *Training in clinical psychology.* New York: Prentice Hall, 1950.
13. AMERICAN PSYCHOLOGICAL ASSOCIATION, DIVISION OF CLINICAL AND ABNORMAL PSYCHOLOGY, COMMITTEE ON PSYCHO-

THERAPY. Report. *Newsletter, Div. clin. abnorm. Psychol.*, 1950, 4, No. 2, Suppl. (Mimeo.)

14. AMERICAN PSYCHOLOGICAL ASSOCIATION, EDUCATION AND TRAINING BOARD. Doctoral training programs in clinical psychology. *Amer. Psychologist*, 1952, 7, 158.

15. AMERICAN PSYCHOLOGICAL ASSOCIATION, POLICY AND PLANNING BOARD. Annual report: 1951. *Amer. Psychologist*, 1951, 6, 531–540.

16. ANDREWS, T. G., & DREESE, M. Military utilization of psychologists during World War II. *Amer. Psychologist*, 1948, 3, 533–538.

17. Anon. Discussion on ethics: a little recent history. *Amer. Psychologist*, 1952, 7, 426–428.

18. BECK, S. J. Rorschach's test in this anniversary year. In L. G. Lowrey (Ed.), *Orthopsychiatry, 1923–1948: retrospect and prospect.* New York: American Orthopsychiatric Association, 1948. Pp. 422–455.

19. BEERS, C. W. *A mind that found itself.* New York: Longmans Green, 1908.

20. BLACK, J. D. A survey of employment in psychology and the place of personnel without the Ph.D. *Amer. Psychologist*, 1949, 4, 38–42.

21. BOBBITT, J. M. Some arguments for a code of ethics. *Amer. Psychologist*, 1952, 7, 428–429.

22. BORING, E. G. Edwin Garrigues Boring. In E. G. Boring, H. S. Langfeld, H. Werner, & R. M. Yerkes (Eds.), *A history of psychology in autobiography.* Vol. IV. Worcester: Clark Univer. Press, 1952. Pp. 27–52.

23. BRILL, A. A. Introduction. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud.* New York: Modern Library, 1938. Pp. 3–32.

24. BRITT, S. H., & MORGAN, JANE D. Military psychologists in World War II. *Amer. Psychologist*, 1946, 1, 423–437.

25. BRONNER, AUGUSTA F., HEALY, W., LOWE, GLADYS M., & SHIMBERG, MYRA E. *A manual of individual mental tests and testing.* Boston: Little Brown, 1927.

26. BROTEMARKLE, R. A. (Ed.) *Clinical psychology: studies in honor of Lightner Witmer.* Philadelphia: Univer. of Pennsylvania Press, 1931.

27. BROTEMARKLE, R. A. Clinical psychology 1896–1946. *J. consult. Psychol.*, 1947, 11, 1–4.

28. BUCK, J. N. The present and future status of the psychologist in the field of mental deficiency. *Amer. J. ment. Def.*, 1949–1950, 54, 225–229.

29. CATTELL, J. M. Mental tests and measurements. *Mind*, 1890, 15, 373–381.

30. CATTELL, J. M. Retrospect: psychology as a profession. *J. consult. Psychol.*, 1937. 1, 1–3; 1946, 10, 289–291.

31. COMBS, A. W. A report of the 1951 licensing effort in New York State. *Amer. Psychologist*, 1951, 6, 541–548.

32. CORIAT, I. H. Some personal reminiscences of psychoanalysis in Boston: an autobiographical note. *Psychoanal. Rev.*, 1945, 32, 1–8.

33. COWAN, E. A. Correspondence. *J. consult. Psychol.*, 1945, 9, 64–65.

34. COWLES, E. Progress during the half century. *Amer. J. Insanity*, 1894, 51, 10–22.

35. DARLEY, J. G., ELLIOTT, R. M., HATHAWAY, S. R., & PATERSON, D. Are psychologists without Ph.D. degrees to be barred from membership in the APA? *Amer. Psychologist*, 1948, 3, 51–53.

36. DENNIS, W., & BORING, E. G. The founding of the APA. *Amer. Psychologist*, 1952, 7, 95–97.

37. DOLL, E. A. (Ed.) *Twenty five years: a memorial volume in commemoration of the 25th anniversary of the Vineland Laboratory, 1906–1931.* (Publ. Ser. 1932.)

38. ELLIS, A. The psychologist in private practice and the good profession. *Amer. Psychologist*, 1952, 7, 129–131.

39. ELLIS, A. (Chm.) Report of the Committee on Private Practice. *Newsletter, Div. clin. abnorm. Psychol.*, 1952, 5, No. 4. (Mimeo.)

40. ENGLISH, H. B. Organization of the American Association of Applied Psychologists. *J. consult. Psychol.*, 1938, 2, 7–16.

41. EYSENCK, H. J. Function and training of the clinical psychologist. *J. ment. Sci.*, 1950, 96, 710–725.

42. FEDERAL SECURITY AGENCY. *National Mental Health Act. Five years of progress, 1946–1951.* Washington, D. C.: 1951. (Mimeo.)

43. FERNBERGER, S. W. The American Psychological Association: a historical summary, 1892–1930. *Psychol. Bull.*, 1932, 29, 1–89.

44. FERNBERGER, S. W. The scientific interests and scientific publications of the members of the American Psychologi-

cal Association. *Psychol. Bull.*, 1938, 35, 261–281.

45. FERNBERGER, S. W. The American Psychological Association, 1892–1942. *Psychol. Rev.*, 1943, 50, 33–60.

46. FINCH, F. H., & ODOROFF, M. E. Employment trends in applied psychology. *J. consult. Psychol.*, 1941, 5, 275–278.

47. FOWERBAUGH, C. C. Legal status of psychologists in Ohio. *J. consult. Psychol.*, 1945, 9, 196–200.

48. FRANK, L. K. Projective methods for the study of personality. *J. Psychol.*, 1939, 8, 389–413.

49. FRANZ, S. I. The present status of psychology in medical education and practice. *J. Amer. med. Ass.*, 1912, 58, 909–911.

50. FRANZ, S. I. Shepard Ivory Franz. In C. Murchison (Ed.), *A history of psychology in autobiography*. Vol. II. Worcester: Clark Univer. Press, 1932. Pp. 89–113.

51. FREUD, S. The origin and development of psychoanalysis. *Amer. J. Psychol.*, 1910, 21, 181–218.

52. FRYER, D. (Chm.) The proposed American Association for Applied and Professional Psychologists. *J. consult. Psychol.*, 1937, 1, 14–16.

53. GALTON, F. *Inquiries into human faculty and its development.* London: J. M. Dent, 1883.

54. GREGG, A. The profession of psychology as seen by a doctor of medicine. *Amer. Psychologist*, 1948, 9, 397–401.

55. GROUP FOR THE ADVANCEMENT OF PSYCHIATRY, COMMITTEE ON CLINICAL PSYCHOLOGY. The relation of clinical psychology to psychiatry. *Amer. J. Orthopsychiat.*, 1950, 22, 346–354.

56. HACKBUSCH, FLORENTINE. Responsibility of the American Association on Mental Deficiency for developing uniform psychological practices in schools for mental defectives. *Amer. J. ment. Def.*, 1940–41, 45, 233–237.

57. HALL, G. S. Laboratory of the McLean Hospital. *Amer. J. Insanity*, 1894, 51, 358–364.

58. HAWLEY, P. R. The importance of clinical psychology in a complete medical program. *J. consult. Psychol.*, 1946, 10, 292–300.

59. HEALY, W., & BRONNER, AUGUSTA F. The child guidance clinic: birth and growth of an idea. In L. G. Lowrey (Ed.), *Orthopsychiatry, 1923–1948: ret-*

rospect and prospect. New York: American Orthopsychiatric Association, 1948. Pp. 14–49.

60. HEISER, K. F. The need for legislation and the complexities of the problem. *Amer. Psychologist*, 1950, 5, 104, 108.

61. HOBBS, N. The development of a code of ethical standards for psychology. *Amer. Psychologist*, 1948, 3, 80–84.

62. HOCH, A. A review of psychological and physiological experiments done in connection with the study of mental diseases. *Psychol. Bull.*, 1904, 1, 241–257.

63. HUNT, W. A. Clinical psychology—science or superstition. *Amer. Psychologist*, 1952, 6, 683–688.

64. JAMES, H. (Ed.) *The letters of William James.* Vol. II. Boston: Atlantic Monthly Press, 1920.

65. JAMES, W. *The principles of psychology.* New York: Holt, 1890.

66. JAMES, W. *The varieties of religious experience.* New York: Longmans Green, 1902.

67. KELLY, G. A. Single level versus legislation for different levels of psychological training and experience. *Amer. Psychologist*, 1950, 5, 109, 111.

68. KRAEPELIN, E. *Lehrbuch der psychiatrie.* Leipzig: Verlag von Johann Ambrosius Barth, 1899.

69. LONGSTAFF, H. P., SPEER, G. S., MCTEER, W., & HARTSON, L. D. A survey of psychologists in four midwestern states. *Amer. Psychologist*, 1950, 5, 422–423.

70. LOUTTIT, C. M. The nature of clinical psychology. *Psychol. Bull.*, 1939, 36, 361–389.

71. LOWREY, L. G. The birth of orthopsychiatry. In L. G. Lowrey (Ed.), *Orthopsychiatry, 1923–1948: retrospect and prospect.* New York: American Orthopsychiatric Association, 1948. Pp. 190–216.

72. MENNINGER, W. C. The relationship of clinical psychology and psychiatry. *Amer. Psychologist*, 1950, 5, 3–15.

73. MERRILL, MAUDE A. Oscillation and progress in clinical psychology. *J. consult. Psychol.*, 1951, 15, 281–289.

74. MEYER, A. G. Stanley Hall, Ph.D., LL.D. *Amer. J. Psychiat.*, 1924–25, 81, 151–153.

75. MEYER, A. AUGUST HOCH, M.D. *Arch. Neurol. Psychiat.*, 1919, 2, 573–576.

76. MEYER, A. Organization of community facilities for prevention, care, and treatment of nervous and mental dis-

eases. *Proc. First Inter. Cong. ment. Hyg.,* 1932, **1**, 237–257.

77. MORRIS, L. *William James: the message of a modern mind.* New York: Scribners, 1950.

78. MORROW, W. R. The development of psychological internship training. *J. consult. Psychol.,* 1946, **10**, 165–183.

79. MURCHISON, C. (Ed.) *The psychological register.* (2 vols.) Worcester: Clark Univer. Press, 1929, 1932.

80. MURPHY, G. *Historical introduction to modern psychology.* (Rev. Ed.) New York: Harcourt Brace, 1949.

81. MURRAY, H. A., *et al. Explorations in personality: a clinical and experimental study of fifty men of college age.* New York: Oxford Univer. Press, 1938.

82. NORSWORTHY, NAOMI. The psychology of mentally deficient children. *Arch. Psychol., N. Y.,* 1906, No. 1.

83. PETERSON, J. *Early conceptions and tests of intelligence.* Yonkers, N. Y.: World Book Co., 1925.

84. PINTNER, R. *Intelligence testing: methods and results.* (New Ed.) New York: Holt, 1931.

85. PINTNER, R., & PATERSON, D. G. *A scale of performance tests.* New York: Appleton-Century, 1917.

86. ROBACK, A. A. *History of American psychology.* New York: Library Publishers, 1952.

87. ROGERS, C. R. Where are we going in clinical psychology? *J. consult. Psychol.,* 1951, **15**, 171–177.

88. ROSENZWEIG, S. Imbalance in clinical psychology. *Amer. Psychologist,* 1950, **5**, 678–680.

89. ROSENZWEIG, S. Balance in clinical psychology: a symposium in correspondence. *Amer. Psychologist,* 1951, **6**, 208–212.

90. SCOTT, W. D. An interpretation of the psycho-analytic method in psychotherapy with a report of a case so treated. *J. abnorm. Psychol.,* 1909, **3**, 371–377.

91. SEASHORE, C. E. *Pioneering in psychology.* Iowa City, Iowa: Univer. of Iowa Press, 1942.

92. SHAKOW, D. One hundred years of American psychiatry: a special review. *Psychol. Bull.,* 1945, **42**, 423–432.

93. SHAKOW, D. Clinical psychology: an evaluation. In L. G. Lowrey (Ed.), *Orthopsychiatry, 1923–1948: retrospect and prospect.* New York: American Orthopsychiatric Association, 1948. Pp. 231–247.

94. SHAKOW, D. Psychology and psychiatry: a dialogue. *Amer. J. Orthopsychiat.,* 1949, **19**, 191–208, 381–396.

95. SHARP, STELLA E. Individual psychology: a study in psychological method. *Amer. J. Psychol.,* 1899, **10**, 329–391.

96. SIDIS, B. Studies in psychopathology. *Boston med. & surg. J.,* 1907, **156**, 321–326, 357–361, 394–398, 432–434, 472–478.

97. SMITH, T. L. The development of psychological clinics in the United States. *Ped. Sem.,* 1914, **21**, 143–153.

98. SPEER, G. S. A survey of psychologists in Illinois. *Amer. Psychologist,* 1950, **5**, 424–426.

99. STEVENSON, G. S., & SMITH, G. *Child guidance clinics: a quarter century of development.* New York: Commonwealth Fund, 1934.

100. SYMONDS, J. P. Ten years of journalism in psychology, 1937–1946; first decade of the Journal of Consulting Psychology. *J. consult. Psychol.,* 1946, **10**, 335–374.

101. TERMAN, L. M. Trails to psychology. In C. Murchison (Ed.), *A history of psychology in autobiography.* Vol. II. Worcester: Clark Univer. Press, 1932. Pp. 297–332.

102. U. S. DEPT. ARMY, OFFICE OF THE SURGEON GENERAL. The U. S. Army's senior psychology student program. *Amer. Psychologist,* 1949, **4**, 424–425.

103. VETERANS ADMINISTRATION. Cooperative training program for clinical psychologists in association with part-time work in VA stations where neuropsychiatric cases are treated. *V. A. Tech. Bull.,* 1948, TB 10A–146.

104. VESTERMARK, S. D. Training and its support under the National Mental Health Act. *Amer. J. Psychiat.,* 1949, **106**, 416–419.

105. WALLIN, J. E. W. *The mental health of the school child.* New Haven: Yale Univer. Press, 1914.

106. WATSON, R. I. The professional status of the clinical psychologist. In R. I. Watson (Ed.), *Readings in the clinical method in psychology.* New York: Harper, 1949. Pp. 29–48.

107. WATSON, R. I. *The clinical method in psychology.* New York: Harper, 1951.

108. WELLS, F. L. *Mental tests in clinical practice.* Yonkers, N. Y.: World Book Co., 1927.

109. WENDT, G. R. Legislation for the general practice of psychology versus legislation for specialties within psychology. *Amer. Psychologist*, 1950, **5**, 107–108.

110. WHIPPLE, G. M. *Manual of physical and mental tests*. Baltimore: Warwick and York, 1910.

111. WHITEHORN, J. C. A century of psychiatric research in America. In J. K. Hall (Ed.), *One hundred years of American psychiatry*. New York: Columbia Univer. Press, 1944. Pp. 167–193.

112. WIENER, D. N. The Minnesota law to certify psychologists. *Amer. Psychologist*, 1951, **6**, 549–553.

113. WISSLER, C. The correlation of mental and physical tests. *Psychol. Monogr.*, 1901, **3**, No. 6 (Whole No. 16).

114. WITMER, L. Clinical psychology. *Psychol. Clin.*, 1907, **1**, 1–9.

115. WITTY, P. S., & THEMAN, VIOLA. The psycho-educational clinic. *J. appl. Psychol.*, 1934, **18**, 369–392.

116. WOLFLE, D. The reorganized American Psychological Association. *Amer. Psychologist*, 1946, **1**, 3–6.

117. WOLFLE, D. Legal control of psychological practice. *Amer. Psychologist*, 1950, **5**, 651–655.

118. WOODWORTH, R. S. *Personal Data Sheet*. Chicago: C. H. Stoelting, 1917.

119. WYATT, F. Clinical psychology and orthopsychiatry. In L. G. Lowrey (Ed.), *Orthopsychiatry, 1923–1948: retrospect and prospect*. New York: American Orthopsychiatric Association, 1948. Pp. 217–230.

120. YERKES, R. M. (Ed.) *Psychological examining in the United States Army*. (Memoirs of the National Academy of Sciences, Vol. 15.) Washington, D. C.: U. S. Govt. Printing Office, 1921.

121. YOUNG, K. The history of mental testing. *Ped. Sem.*, 1924, **31**, 1–48.

PSYCHOLOGY IN ITALY

HENRYK MISIAK
Fordham University

AND

VIRGINIA M. STAUDT
Hunter College

Italy very early took cognizance of scientific psychology and joined other countries in accepting and developing the new science (**2, 6, 9, 28**). Although she made significant contributions to the study of human behavior, her influence in the history of psychology was never very profound.

The attainments of Italian psychologists have not been equal in importance to those made by Italian scientists in allied areas like psychiatry and neurology. The contributions of Italy in these latter fields have attracted worldwide attention (**16**). The first homes for the insane were established in Italy as early as the 14th century, and the psychiatric hospital Santa Maria della Pietà in Rome, founded in 1548 and still functioning, is the oldest in the country. The reforms in the treatment of patients, effected by Chiarugi (1759–1820), actually antedated those of the French physician Pinel (1745–1826), who is credited with being the first to revolutionize the treatment of the insane. Chiarugi was also the author of the first Italian text in psychiatry. Music therapy, occupational therapy, and even psychodrama were used by Miraglia in Naples in the middle of the 19th century. In Italy psychiatry has always had a neurological orientation. Therefore, it is not surprising that in recent times in Italy electroshock (Cerletti, Bini) and acetylcholine induced shock (Fiamberti) should have originated. Supraorbital lobotomy in psychosurgery (Fiamberti) also had its origins in Italy. In neuroanatomy and neurology such names as Golgi, Luciani, Marchi, Bianchi and their discoveries are well known.

BEGINNINGS AND PIONEERS

Materialistic positivism had a strong hold on early Italian psychology mainly because of the powerful and enduring influence of Roberto Ardigò, professor at Padua University, who in 1870 published *La Psicologia come Scienza Positiva* and in 1898 *Unità della Coscienza* in which he identified all mental life with cerebral physiology. In the idealistic reaction against positivism, psychology, which was considered an offspring of this system and a pseudophilosophy, was opposed together with positivism. As a result psychology was excluded from, or greatly reduced in, the school curricula. The early strongholds of psychology were the universities at Rome, Florence, and Turin.

The first Italian psychologist was Giuseppe Sergi (1841–1936), professor at the University of Rome. The year of publication of his book *Principi di Psicologia*, 1873, the same year as that of Wundt's first volume of *Grundzüge der physiologischen Psychologie*, is regarded as the birth date of Italian psychology. In 1885 the first Italian psychological laboratory was established by him as a section of the Institute of Anthropology in Rome. Although Sergi wanted psychology to be independent of philosophy, his own philosophical credo, strong positivism, permeated his views in psychology.

347

A different philosophy and a different philosophical trend were represented by another pioneer of psychology in Italy, Francesco de Sarlo (1864–1937), a devoted disciple of Brentano, at the University of Florence (4). Through his efforts the first institute of psychology was opened in Florence in 1903. De Sarlo was a philosopher, psychiatrist, surgeon, and psychologist. His views in psychiatry anticipated the psychobiological concept, later to become so popular, but without violation of the traditional dualism. He received his training from the director of the Psychiatric Hospital in Reggio Emilia, Augusto Tamburini, a man who was greatly interested in psychology and influenced its development in Italy through his students, such as De Sarlo, G. Buccola, and C. G. Ferrari. It was Tamburini who first popularized the concept of mental hygiene among the Italians. Buccola (1854–1885) was the author of *La Legge del Tempo nei Fenomeni del Pensiero* (1883), a work which aroused a current of sympathy for psychological research in Italy. Ferrari (1869–1932) contributed substantially to the development of psychology: with Tamburini he founded a psychological laboratory in Reggio Emilia in 1896, the magazine *Rivista di Psicologia*, and in 1901 he translated W. James's *Principles of Psychology*, an event which has been acknowledged as the turning point in the history of Italian psychology (9, 10).

In Turin psychological problems were introduced into the field of physiology by Angelo Mosso (1846–1910) and into criminal anthropology by Cesare Lombroso (1835–1909). The former is well known to psychologists for his ergograph and for his pioneering research on work and fatigue. His books, *La Paura* (1884) and *La Fatica* (1891), were translated into many languages, including English. Mosso became famous for his *L'Uomo Delinquente* (1876) and was hailed as the forerunner of constitutional psychology. In general, Italians, and especially such scientists as Di Giovanni, Viola, and Pende, have contributed to constitutional psychology by their anthropometric techniques and their vigorous efforts to establish the relationship between physique and personality. Although Lombroso did not develop any system, his observations and his views aroused a great deal of controversy and prompted much research. In later years he directed his attention to parapsychology. In 1895 Mosso's laboratory was turned over to Friedrich Kiesow (15) (1858–1940), a pupil and former assistant of Wundt's and a friend of Külpe's. Kiesow transplanted from Leipzig to Turin not only the knowledge of experimental techniques, but also an enthusiasm for experimentation in psychology. For many years, long before others opened laboratories and began research in other places, he was the greatest experimentalist in Italy. His own work was mainly in the field of sensation, in taste and touch especially. Among Kiesow's students were Gemelli and Ponzo. Another experimentalist imported from abroad was Vittorio Benussi (1878–1927), who, although born in Italy, lived in Austria for many years and became the most outstanding experimentalist there. After World War I he went to Padua, taught psychology at the university there, and opened a laboratory. His studies on perception, respiration, suggestibility, and hypnosis received recognition in Italy and abroad.

A special place among Italy's

pioneers in psychology is held by Sante de Sanctis (1863–1935), an eminent figure not only in psychology, but also in psychiatry (1, 7). Through his enthusiasm for the new science, his versatile activities, his writings, and his personal charm he left a deep impression on Italian psychology. He succeeded his teacher, Sergi, as head of the school of Rome and became the first graduate teacher of experimental psychology in Italy. While practically all the fields of theoretical and applied psychology were explored by De Sanctis, his best energies were devoted to child study on the one hand and to the understanding and helping of the mentally deficient and the abnormal on the other hand. His publications were numerous, original, and influential. Interested in sleep and dreams, De Sanctis published several studies on the subject in 1896 and a large monograph, *I Sogni* (Dreams), in 1899 which preceded that of Freud. Other important writings include *La Mimica del Pensiero* (1904), *Psicologia Sperimentale* (Vol. 1, 1929; Vol. II, 1930), and *La Conversione Religiosa* (1924) (translated into English, *Religious Conversions: a Bio-Psychological Study*, [1927]). He professed recognition of and respect for religious faith. "I had dealings with materialists, positivists, rationalists," he said, "but they all failed to inculcate into me their philosophical convictions," and he said further that no one succeeded in shaking his belief in finality in nature and the universe (7).

PROGRESS AFTER 1905

In 1905 a significant event took place in Italy, the 5th International Congress of Psychology in Rome, under the presidency of Sergi, one of the most provocative congresses of psychology ever held. At this congress the first scale of intelligence, the Binet-Simon, and W. James's paper, "La Conscience Existe-t-Elle?" were given prominence. In general the congress found psychology in Italy firmly established. The work of Italian psychologists and of their laboratories was already well recognized. There was a lively interest in child study. The new studies in this field conducted in Britain, Germany, and France were matched by the excellent medical and biological studies of the child in Italy. Melzi, De Sanctis, and Ferrari had already distinguished themselves in that field, and Maria Montessori was about to launch her new educational movement. There were two dominant currents of psychological thought, one stemming from Wundt, the other from Münsterberg (6).

There was one significant practical outcome of the congress for Italy: the next year the ministry of education formally instituted three autonomous chairs of psychology on the university level, in Rome, Turin, and Naples, given respectively to De Sanctis, Kiesow, and Colucci. De Sanctis was later succeeded by Mario Ponzo, Kiesow by Alessandro Gatti, and the latter after his premature death by Angiola Massucco Costa. Galdo succeeded Colucci.

The other centers of psychological research and teaching have been Florence, Padua, Milan, and Genoa. In Florence the institute of psychology was headed by De Sarlo, then by Enzo Bonaventura, and since 1938 by Alberto Marzi. The chair of psychology in Padua was occupied by Benussi, and from 1927 to 1948 by Cesare Musatti. In 1948 when Musatti accepted the chair of psychology at the State University in Milan (founded three years after the

Catholic University in Milan, in 1924) he was succeeded at Padua by Fabio Metelli. Musatti, who has been one of the few exponents of psychoanalysis, is the author of a two-volume work on the subject. In the other university of Milan, the Catholic University of the Sacred Heart, Agostino Gemelli has been the head of the psychology department. In Genoa Giuseppe Vibone has taught psychology at the University. There the faculty of medicine has also instituted a chair of psychology, of which Amedeo Dalla Volta is head.

In World War I, during the political upheaval in Italy in the nineteen thirties, and more recently in World War II, psychology suffered a great blow. In spite of these hindrances progress has continued, however, marked among other events by the creation of the Italian Psychological Association and by the state decree in 1935, making psychology a required subject for a degree in philosophy and complementary for a degree in pedagogy, medicine, and law (19, 20, 27). The opening in 1940 of an experimental center for applied psychology in the National Research Council, with Ferruccio Banissoni as director, is to be mentioned particularly for it gave a new impetus to psychological research (3).

TRENDS IN ITALIAN PSYCHOLOGY

Following the period of interest in psychophysics, perception, and child study, so prominent prior to 1920, the emphasis in contemporary Italian psychology shifted to applied fields. A number of private and public centers for psychotechnical research were opened and almost all leading psychologists were engaged at some time or other in applied psychology. Aptitude testing and vocational guidance received special prominence. There were strong efforts to extend psychology to education through the development of new methods of education based on the findings of child psychology, by instituting projects for school reform, by training teachers in psychology, and by similar means. Child psychology has been fostered at the University of Florence. The psychology of personality, or characterology as it is called there, has attracted many adherents in Italy. Psychoanalysis, however, as a doctrine and method has had very few followers and is taught at very few universities. The religious, cultural, and social traditions of Italy have not allowed psychoanalysis to enjoy the favorable position which it has had in other countries; in fact, there have always been more opponents than defenders of psychoanalysis (23).

LEADING CONTEMPORARY SCHOOLS AND PSYCHOLOGISTS

Let us now review the important centers of psychology and acquaint ourselves with the leading psychologists of Italy. It is obvious that it is the ability and talents of the men at these universities that determine the role and influence of these schools on Italian psychology. Thus it will be mainly the psychologists themselves on whom we shall focus our attention. Following a historical sequence, we shall speak of Ponzo and the University of Rome, Marzi and Metelli and the University of Florence, Bari and Padua, Gemelli and the Catholic University of Milan, and Banissoni and the National Institute of Psychology. Then we shall mention the new centers of psychology at the Gregorian University and at the Salesian University. The most prominent living psychologists from the point of view of the volume and

extent of work and influence are Gemelli, Ponzo, Marzi, and Metelli.

University of Rome

As mentioned before, the teaching and the laboratory of psychology were begun here by Sergi. Under the subsequent direction of Sante de Sanctis, student and collaborator of Sergi, the school of Rome flourished and gained the prestige which it still has under the leadership of Mario Ponzo, successor of De Sanctis. Ponzo has become the most representative psychologist of Italy. Born in 1882 in Milan, he studied in Turin, where he received the doctorate in medicine, and then taught psychology at the University of Turin from 1905 to 1931. He was Kiesow's pupil and then his colleague and collaborator. In 1931 he went to Rome and has been there ever since as a professor of psychology and head of the Institute of Psychology. Ponzo is the president of the Italian Psychological Association, which after many years of inactivity was revived in 1951. He is also a member of the Executive Committee of the International Union of Scientific Psychology. An indefatigable worker and an unusually prolific writer, Ponzo has for a long time been associated with psychophysical problems, especially touch and taste. Later he extended his research to other fields of psychology and in the last two decades he has devoted himself particularly to applied psychology. With a profound conviction in the usefulness and possibilities of applied psychology, Ponzo has strenuously fought for its recognition in education, industry, vocational guidance, and personnel selection. His own studies are valuable contributions to aptitude testing and professional selection. According to Ponzo, vocational guidance should not be based solely on the assessment of aptitudes or intelligence, but on the general personality characteristics. Personality, character, and the natural inclinations of the individual were frequently stressed by Ponzo as the primary consideration, whether in vocational guidance, personnel selection, or accident prevention. The recognition and growth of applied psychology and the emphasis on the problems of applied psychology in contemporary Italian psychology were undoubtedly to a great degree the result of Ponzo's writings and activities, which also produced practical effects in various spheres of the national life of Italy.

Upon succeeding De Sanctis, Mario Ponzo dedicated himself to organizing the Institute of Psychology of the University of Rome, which he unfortunately saw ruined during the war as a result of the bombings. Since World War II, owing to his efforts and zeal, the Institute has again revived its activity and offers much promise of promoting psychological studies. Each member of the staff has his particular area of interest. Ponzo himself at present is interested in group psychology and vocational guidance. Leandro Canestrelli, assistant director at the Institute, works primarily in the areas of measurement of psychomotor skills and general experimental psychology and has distinguished himself by original studies of psychomotor activity and the use of the photocyclographic method photographing the trajectories covered during movement in the study of voluntary acts.

The second assistant, Ernesto Valentini, is an animal psychologist. Gigliola Sbordoni specializes in the field of psychometrics, child psychology, and audiovisual aids in teach-

ing. The remaining three assistants serve without compensation. Franco, Ferracuti, and Sbrana work in clinical psychology, Franco being especially interested in the psychological problems of senescence and of cancer patients. A clinical psychologist in the Italian Air Force, Nicolo Numeroso, is working on personnel selection and is also interested in psychodrama. In addition to their experimental work, Ponzo and his staff also teach rather large psychology classes, conduct a child guidance clinic and a school for social workers, have a private clinical practice, and do some industrial counseling.

The laboratory at Rome itself is well equipped as far as the older standard apparatus and space are concerned. Experienced machinists build special apparatus for the experimentation and there are excellent facilities for constructing equipment. Most of the research is done primarily in the field of perception. An account of the work accomplished during the last five years in the laboratory is presented in the Proceedings of the IXth Congress of Italian Psychology (Nov. 1–4, 1951). There are lecture halls, which are very suitable for laboratory demonstrations, and visual aids including movie projection equipment. The library is well provided with Italian, French, and German journals, but the American and English journals are noticeably lacking.

University of Florence

De Sarlo, founder and director of this first institute of psychology in Italy, was succeeded by Enzo Bonaventura, who demonstrated the significance of empirical factors in the genesis of perception by his research on perception, especially on the perception of time and space. Bona-

ventura published several volumes; the most successful was his work on psychoanalysis, La Psicoanalisi (last edition in 1950). Because of anti-Semitic attitudes in Italy at that time he was forced to relinquish his post as director of the institute and was finally compelled to leave Italy. Up to the time of his tragic death in 1948, which occurred during an Arab attack when he was giving aid during the war in Palestine, Bonaventura held the seat of psychology at the University of Jerusalem.

Bonaventura's successor in 1938 was Alberto Marzi, a prominent figure of growing influence in Italian psychology today. His research has included a variety of fields such as attention, eidetic imagery, the nature of intelligence of deafmutes, and psychotechnical problems. Marzi was associate professor at the University of Florence until 1948 and then after taking the requisite government examinations he became professor. Having received a Rockefeller grant, he visited the United States for five months. Following this visit he proceeded to the National Institute of Psychology in London and then to Paris where he studied child psychology with Henri Wallon. When Marzi returned to Italy he became the head of psychology at the University of Bari in southern Italy, the fourth most important Italian university. Thus Marzi continues to hold academic appointments at Florence and at Bari, a fact that indicates the shortage of senior psychologists in Italy (14). There are also other psychologists with dual appointments, such as Banissoni and Metelli (11). The modest Laboratory of Industrial Psychology in Florence, directed by Marzi, is financed by the city and it provides testing services to industry and to the school system.

Marzi is assisted by Maria Louisa Falorni, who directs the practical work of four young psychologists and seven social workers. His other activities include the direction of the Center for Studies of Accidents which was established in 1951 and which has a staff of five psychologists, all graduates of the University of Florence, and one social worker.

In addition to Marzi's teaching duties in Florence he has been translating the Wechsler-Bellevue test for children with the assistance of Lydia de Rita and he has been studying projective techniques with Giovanni Battista Guarini. At Bari, however, Marzi's major interest has been the psychological study of social classes. He has been engaging in a study of some seven hundred cave dwellers from the nearby villages of Matera and Lucania.

Despite his many teaching and research interests and his travels between Florence and Bari, Marzi finds time to edit *Rivista di Psicologia*, the bulletin of the Italian Psychological Association, of which he is general secretary. He is also a foreign affiliate of the American Psychological Association. Marzi's work at Florence has been published in two volumes entitled *Studie Ricerche Università delle Firenze 1938–1947* (Vol. I) and *1947–1949* (Vol. II). He has prepared a third volume on *Studie Ricerche Instituto di Bari* (14). At present he is writing a book on the psychology of work and he plans to translate the next edition of Viteles' *Industrial Psychology*.

Fabio Metelli presides over the laboratory which is operated under the auspices of the department of philosophy of the University of Florence, where he is assistant professor. Strangely enough, however, Metelli's real teaching appointment is at the University of Padua, and he spends three days each week in Florence and three at Padua. Metelli considers the continuation of the laboratory at Florence a distinct service to psychology (11). He attributes the lack of adequate support for the laboratory, which is operated on an annual budget of about eighty dollars, to the fact that Italian philosophers are too idealistic and therefore are not very enthusiastic about science. His own research and writings have covered such areas as dreams, testimony, and the perception of movement.

At the University of Padua one course is given in psychology which varies each year—individual differences, memory, perception. Consequently, in order to cover the field the student must continue to take this course for several years. He must also get a good deal of his background in psychology from his own reading. Demonstrations are provided in the course, but there is no opportunity for laboratory practice. The laboratory at Padua, located in a 15th century building, has very limited space, apparatus, and library facilities. Much of the research done is in the field of configural perception after the tradition of Benussi (11).

Catholic University in Milan

The best equipped laboratory, not only in Italy but probably in all continental Europe, is at the Catholic University in Milan (14). It was established and has been directed by a Catholic priest, Agostino Gemelli, the most prominent contemporary Italian psychologist, whose influence in this Catholic country (99.6% of the inhabitants are Catholic) has been very significant with respect to the attitude of Catholics towards

psychology. This influence marks a new era in Italian psychology.

Gemelli (5, 12, 18, 20) was born in 1878 and he studied medicine at the University of Pavia. After his doctorate in medicine and surgery he remained at the university and continued research as an assistant to Professor Camillo Golgi (1844–1926), famous anatomist and physician, Nobel prize winner. During this period he also studied philosophy and was engaged in social and political activities. Abandoning religion, he embraced materialism and Marxism, but gradually became dissatisfied and disillusioned with these systems. Finally he returned to the Church, entered the Franciscan order, and was ordained a priest in 1906. From 1907 to 1911 Gemelli studied biology, physiology, and philosophy at various universities in Bonn, Frankfort on Main, Munich, Cologne, Vienna, Louvain, Amsterdam, and Paris. In 1911 he completed his doctorate in philosophy at the University of Louvain.

The psychologist who gave Gemelli his initial training in experimental psychology was Friedrich Kiesow, who from 1899 was on the faculty of the University of Turin and who recommended Gemelli to his friend Oswald Külpe. Gemelli went to Külpe at Bonn and then with him to Munich, working in his laboratories. Gemelli has always recognized the profound impression that Külpe made on him, and through the years he has continued to regard himself as a "student of Külpe" (12). The interrupted association with Kiesow was renewed again and continued afterwards for many years when Gemelli accepted the appointment by the Italian government to teach psychology at the University of Turin in 1914. With Kiesow he founded the *Archivio di Psicologia, Neurologia e Psichiatria*.

During World War I Gemelli served as a military chaplain, a physician, and a psychologist. In the latter capacity he became known for his work in the selection of pilots. After the war his efforts were directed towards the founding of a Catholic university which was officially opened in 1921 in Milan under the name "The Catholic University of the Sacred Heart in Milan." From its founding Gemelli has been the rector of the university and has done a great deal to promote the teaching of psychology and psychological research. In the subsequent years Gemelli has conducted a large number of original researches and published many books and articles. His bibliography comprises several hundred items, among them such books as *La Psicologia del Pilota di Velivolo* (1942), *La Psicologia Applicata all'industria* (1944), *L'Orientamento Professionale dei Giovani nelle Scuole* (2nd Rev. Ed., 1947), *Introduzione alla Psicologia* (2nd Rev. and enlarged Ed., 1949). These are simply the latest works.

From a perusal of Gemelli's writings, taken in chronological order, it is evident that his initial studies were in the most difficult area of psychology, the so-called higher mental processes, thought and will, an obvious outcome of his association with Külpe and Michotte. Then when he was with Kiesow in Turin the major areas of his work were psychophysics, sensation, and perception. At Milan in his own laboratory the most original was his research on language. In the meantime his interest in applied psychology grew, a field to which he made appreciable contributions, particularly in aviation, industry, education, and criminology.

As a spokesman for, and interpreter of, the new psychology for Catholics, Gemelli was forceful and effective. He clarified for Catholics the status of psychology as a science, defined its relationship to philosophy, biology, and other disciplines, and most of all devoted himself to experimental research, contributing to psychology, adding something of his own to its treasure. He dispelled the last shreds of suspicion and doubt about the new psychology, especially in Italy. His authority and prestige carried weight. After all, he was a physician, a psychologist, a theologian, a priest, and a monk all in one. Moreover, he was a rector of a Catholic university and the president of the Pontifical Academy of Sciences. His own example spoke louder than his words. That there can be harmony between science and faith, between modern progress and old tradition, between psychology and philosophy, and between speculation and experiment, his writings explained and his example proved.

Biological orientation has been a characteristic feature of Gemelli's psychology. Some Catholic psychologists have the tendency, perhaps, to let the biological counterpart of man be overshadowed by the spiritual or mental, to divide man too much into the physical and psychical, to minimize the bodily component in psychological processes. Gemelli has avoided that. Whether in emotion, or perception, or delinquency, or in other fields, he has been fully aware of organismic forces. The biological orientation which also made him bring psychology closer to biology than to philosophy resulted from both his medical training and his consistent application of the Aristotelian hylomorphic doctrine of man to his work and theories. What-

ever aspect of man he considered or examined, he never lost sight of the unity of man.

In the Catholic University of the Sacred Heart in Milan psychology received special attention. The systematic teaching of experimental psychology was included in the program of the university and a special laboratory for training and psychological research was opened. Gemelli assumed the role of professor of psychology and director of the laboratory (18).

The work in the Laboratory of Psychology has been very fruitful. Its productivity is demonstrated in the numerous volumes of a special series of monographs, *Contributi del Laboratorio di Psicologia*, which contains original studies completed by Gemelli and his collaborators. The areas of research at the Laboratory are varied, but the best studies have been done in the field of perception, personality, and language. In perception the specific problems investigated included: individual differences, localization of sound, perception of depth, and spatial perception. The theoretical stand assumed on the basis of these studies places the school of Milan between the elementaristic view along the lines of Wundtian tradition and that of the Gestalt school. The school maintains that even though we normally have a unification of sensory data, there is also the possibility of perceiving isolated elements without constituting them into a whole. In the field of motor activity with human and animal subjects it was demonstrated how a task unifies and renders intelligible various phases of activity. The most original and brilliant was the study of language based on the electroacoustic recordings which gave new insight into the

understanding of language. Extensive research was done in electroencephalography, the Laboratory being the first to begin this type of research in Italy. During World War II the Laboratory engaged in special work for both the war effort and the city of Milan. Other areas include psychometrics, child psychology, social psychology, industrial psychology, and gerontology.

During the war the buildings of the Laboratory were completely destroyed but the equipment was saved. The total reconstruction took several years and only in 1950 could the Laboratory resume its work fully. Gemelli now has nine assistants. Five of them have a doctorate in medicine, three in pedagogy, and one in philosophy. There are about twenty graduate students working on different problems for their dissertations.

A prominent figure at the University of the Sacred Heart has been a priest, Giorgio Zunini (born in 1903), professor of psychology there and Gemelli's closest collaborator. He began his scientific career at the University of Pavia. From geology and paleontology he turned to biology and then to comparative psychology and systematic psychology. Among his publications in comparative psychology are excellent and original studies on learning in fish. In his book, *Animali e Uomo; Visti da uno Psicologo* (1947), he shows the gradual development of behavior from fish to man and the qualitative differences existing between man and animals. The systematic works include *Introduzione alla Psicologia* (2nd Ed., 1949) published with Gemelli and *Psicologia* (2nd Ed., 1950), a brief history of psychology and exposition of psychological systems. Both books filled a need that was sorely felt in Italian psychological literature.

The University of the Sacred Heart has been the place where many psychologists have begun their professional careers. A number of Italians as well as foreigners received their training and degrees there and carried out their studies using the facilities of the Psychological Laboratory. Among the Italians there were several who distinguished themselves. Arcangelo Galli, a pupil of Michotte at Louvain and of Wertheimer and Gell at Frankfort, did several experimental studies at Louvain (1914, 1924). Even before the founding of the Catholic University he published a study with Gemelli and Tessier on the perception of the position of the body. During the many years of his association with the Laboratory he made several contributions mainly in the field of perception as well as some in applied psychology in collaboration with Gemelli. Alessandro Gatti (25), author of articles on psychotechnical and perceptual problems and of a book, *Le Massime e i Caratteri* (1934), went to the University of Turin, where he founded a center of studies on work, the only center of this type in Italy. In 1928–1929 Gatti visited the United States as a Laura Rockefeller fellow. He died in Turin at the height of his professional activity. Giuseppina Pastori, now director of the Laboratory of Biology at the Catholic University, collaborated with Gemelli on his electroacoustic study of language. G. Sacerdote became director of the Institute of Electroacoustics in Rome. C. Trabattoni worked in electroencephalography, and L. Ancona studied stereoscopic "aftereffects."

The students at the University and investigators at the Laboratory have come from many foreign countries:

from Austria, Hubert Rohracher, now professor at the University of Vienna and the most prominent experimentalist in Austria; from Bulgaria, B. Raduseff; from Finland, A. Pentilla; from Yugoslavia, A. Terstenjak, P. Gubernia, P. Matko; from Lithuania, A. Sidlauskas, now assistant professor at the University of Ottawa, Canada; from Rumania, A. Manoil, who published a large volume on the Milan School (18), and who is now chairman of the department of psychology at Park College, Parkville, Missouri; from China, P. Siao Sci Wi. It is evident from this enumeration that the influence of the Catholic University in Milan has been very widespread.

The National Institute of Psychology in Rome

The National Institute of Psychology had its origins in the Experimental Center of Applied Psychology of the National Research Council in 1940. Later in 1946 it was called the Center of Psychological Studies and then on January 1, 1950 its title was changed to the National Institute of Psychology. There are fourteen laboratories affiliated with the Institute, and of this number ten are situated in the university cities of Bari, Bologna, Catania, Florence, Genoa, Messina, Naples, Padua, Trieste, and Turin.

The Institute and its collaborating laboratories attempt to furnish scientific counsel and guidance to public and governmental offices on a variety of psychological problems, such as the selection of scientific and civil service personnel, air force, military, naval, and police personnel selection, vocational guidance for students at the age of eleven years for classical and higher education, industrial counseling, and accident prevention. As a result of these studies and of the collection of a tremendous amount of data, it is possible to make population analyses and comparative studies of the inhabitants of the various regions of Italy.

The psychological activities of the Institute have also extended to other areas. New tests are being devised; national standardizations are being developed; foreign tests, such as the Terman-Merrill, are being translated, adapted, and standardized. A variety of research investigations are being conducted on such topics as the interrelationships of personality and environment, characterology, and prejudice. In 1951 at the 13th International Congress of Psychology in Stockholm and at the 10th International Congress of Psychotechnics at Gothenberg, Banissoni gave an account of some of the Institute's psychosocial studies. Through the Institute's research, Professor Banissoni hoped to emphasize the usefulness of applied psychology and thus to encourage the universities to foster and improve the teaching of general and experimental psychology which for years had been poorly done in the departments of philosophy. The Center of Technico-Scientific Documentation of the National Research Council receives the reports of the National Institute of Psychology in the fields of general and applied psychology and publishes notices of them in its *Index of Scientific and Technical Periodicals.*

In addition to all these activities the director and staff members of the Institute offer training courses for the Institute's personnel and they also give courses to the Institute's external collaborators. For the most part, these courses are brief, conducted over a period from a few weeks to three months and covering a fairly

wide range of topics: statistics in psychological research, educational psychology, social psychology, applications of the Szondi projective test, and psychosomatic medicine. Such lectures and courses have been held in Florence, Trieste, Civitavecchia, Udine, and Salerno, as well as in Rome.

The organizer of the National Institute of Psychology was Ferruccio Banissoni. Born in Trieste in 1888, he studied in Vienna for four years and in Rome for two years. In 1921 he received the doctorate in medicine. For many years he was a professor at the University of Rome and he also lectured and gave courses at other institutions. His work, broad in scope and marked by originality and progressiveness, contributed to many fields of experimental, applied, and clinical psychology. With his training and research in medicine he was able to treat with authority the problems that were on the borderline of psychology and medicine, as for example his work on the application of the electrocardiogram in experimental psychology. Together with Ach, Michotte, De Sanctis, Lindworsky, and Abramowski, Banissoni is one of the few psychologists who studied experimentally one of the most difficult problems of psychology, the will, and wrote on the subject (*Contributo alla Psicologia Sperimentale della Volontà*, 1926). He and an outstanding Polish psychologist, M. Dybowski, had developed a plan for research on will problems in 1938, a plan which the latter carried out later in Poland (8). As head of the psychological section of the National Research Council, Banissoni inspired and directed a great many studies in applied psychology. The psychology of work, aptitude testing, personnel selection,

and a new branch of applied psychology called filmology have been the areas of his most fruitful research. In contemporary Italy it was Banissoni who raised psychology to new heights, and so his death in 1952 was a real loss to psychology.

The Gregorian University in Rome

For several centuries this venerable institution has been a place where clerical students from all over the world received their philosophical and theological training. The school is administered by the Jesuit Fathers and all the professors are Jesuits from various countries throughout the world. Since the University is primarily devoted to teaching seminarians and priests, it emphasizes philosophy and theology in its curriculum. Thus far Gregorian University has not offered many courses in scientific, social, or pastoral psychology and it has had no facilities for laboratory work, in spite of the presence of such excellent teachers of psychology as Joseph Fröbes, Johannes Lindworsky, Alexander Willwoll, and Paul Siwek. In recent years, however, a more serious effort has been made toward a better integrated program in psychology by a new professor, André Godin (born in 1915, Belgium). After working in criminal psychopathology under the direction of Professor De Greeff (Louvain), he received his doctorate in philosophy in Brussels, Belgium (1942), where he later stayed at the "Centre de Consultations Médico-Psychologiques," thus acquiring both good theoretical foundations and extensive practice in counseling and guidance. In 1949 he was named assistant professor of educational psychology at the Gregorian University. Feeling the need for additional

knowledge of experimental methods and statistics in psychology, he came to the United States and studied at Fordham University where he received a master's degree in psychology. During this period of study he visited many centers of psychological research in America. Upon returning to Rome he took with him not only a thorough knowledge of American psychology, but in addition some of the modern psychological apparatus, psychological films, and books for his work at the university there. Besides publishing articles in various magazines here and abroad he has delivered papers at two international congresses of psychology: in 1948 at Edinburgh and in 1951 at Stockholm. Godin's main interest lies in clinical psychology, particularly in psychoanalytic doctrine and method. The major task that he has posed for his scientific work is to bridge the psychoanalytic dynamics with moral and religious values. In the summer of 1951 Godin helped to organize an international conference of Jesuit psychologists, the first of its kind, which met in England.

The Institute of Experimental Psychology of the Pontifical Salesian University in Turin

The creation of this Institute by the Salesian Fathers in 1938 was a significant stride in the progress of experimental psychology among Catholics in Italy (17). The Institute has classrooms, a library, and a spacious laboratory that is well equipped with modern apparatus and is recognized as one of the best in the country. The courses offered cover a wide range of psychology: general experimental psychology, genetic psychology, characterology and biotypology, social psychology, psychological testing, ab-

normal psychology, industrial psychology, criminal psychology; and psychology of religion as preparation for teachers of religion. Seminars are also held. The main purpose of the Institute is to offer the latest findings in all fields of experimental psychology to the students who are engaged in philosophical studies. In original research the results of the Institute have not been impressive, but some interesting studies have been done with respect to vocational guidance and aptitude testing. During the war the Institute concerned itself with the psychological effects brought about by shock. Wounded soldiers and civilians who had been or were suffering from shock due to the bombing were studied and a program of rehabilitation for them was devised. In 1942 the Institute itself was bombed.

The director of this Institute has been a Salesian priest, Giacomo Lorenzini (born in 1909). Before becoming the director of the Institute of Experimental Psychology he taught philosophy at the Salesian Philosophical Institute of St. John Bosco in Turin. At the former Institute he has given courses in experimental and educational psychology. His main interest is in the psychology of adolescence with particular reference to its religious, social, moral, and esthetic aspects. Several of Lorenzini's publications deal with educational problems and personality typology. One is a systematic textbook, *Corso di Psicologia* (1948).

EVALUATION

In comparison to other countries psychology in Italy did not develop quickly. Experimental psychology was not as readily incorporated into the programs of institutions of higher

education. The appreciation of psychology and its influence on cultural and social life and on education have been very limited. At the meeting of the International Bureau of Education in Geneva in 1937, Italy was the only nation out of forty-two participating nations which did not agree to include psychology as a required course in the training of teachers. The chairs of psychology were instituted late and remained few. The number of psychological laboratories was small. In general, Catholic schools and institutions kept aloof. There was a strong resistance to psychology as an independent science. The recognition in wider circles of experimental psychology as a separate science, and not simply as a part of or as an adjunct to philosophy or physiology, came much later, only in the 1920's.

The reasons for the slow development of psychology and its limited influence in Italy have been multiple. They are of an economic, social, religious, and even a political nature. The proper evaluation of all these factors is not a simple task, but it is immediately obvious that the chief difficulty for psychology arose at the very outset when it was linked with a positivistic philosophy. Psychology used to be regarded as a product of that philosophy, and was practically identified with it. Thus the succeeding philosophical current, antagonistic to positivism and seeking to banish it, also rejected experimental psychology, its alleged product. Moreover the protagonists of the new psychology were often professed agnostics or outright materialists, not infrequently open enemies of religion who used psychology in their fight against religion. This state of affairs could not create a favorable atmosphere for psychology in Catholic Italy, and instead of finding favor psychology met with opposition, suspicion, or indifference. The picture was altered only when a Franciscan priest, Agostino Gemelli, showed that experimental psychology is a separate autonomous science and per se does not contradict any philosophical or religious principles. The founding of the psychological laboratory at the Catholic University in Milan and the prestige of Gemelli overcame the suspicion and opposition toward experimental psychology among Catholics. The role of Gemelli and of the Catholic University in Milan in gaining acceptance for psychology by Catholics cannot be emphasized strongly enough. But one must not overlook the important influence of those Catholic psychologists whose prestige in Italy has been very great, namely Ponzo and Banissoni.

In spite of all these gains, psychology in Italy still has not attained the place and role that it should enjoy. It is still relegated for the most part to the faculty of philosophy as a complementary discipline. It attracts a very small following in university teaching. There are very few individuals who dedicate themselves to psychology because there is little possibility of remunerative work for a qualified psychologist. For the student who writes his thesis in psychology usually the only occupational prospect available is teaching philosophy (19, 20). Nevertheless psychology in Italy now is slowly but surely acquiring the scientific dignity and social prominence that it merits and has gained in other countries. As for the Catholic scientists and the Catholic institutions of learning, they are participating more and more in Italian psychology and their contributions are steadily growing in volume and quality.

REFERENCES

1. APPICCIAFUOCO, R. La psicologia speri-mentale di Sante de Sanctis. Roma: Orsa Maggiore, 1946.

2. BANISSONI, F. Psicologia sperimentale. In Various, Un secolo di progresso scien-tifico italiano, 1839–1939. Roma: S.T.P.S., 1939, Vol. IV. Pp. 377–427.

3. BANISSONI, F. Instituto Nazionale di Psicologia. Ricerca Scient., 1950, 20, 1625–1629.

4. BENUSSI, V. Die Psychologie in Italien. Arch. gesam. Psychol., 1906, 7, Litera-turbericht, 141–180.

5. BOTAZZI, F. Agostino Gemelli et ses études psychologiques. Scientia, Mi-lano, 1940, 67, 115–124.

6. CHIABRA, G. The tendencies of experi-mental psychology in Italy. Amer. J. Psychol., 1904, 15, 515–525.

7. DE SANCTIS, SANTE. Autobiography. In C. Murchison (Ed.), History of psy-chology in autobiography. Vol. 3. Worcester: Clark Univer. Press, 1936. Pp. 83–120.

8. DYBOWSKI, M. Dzialanie woli (The ac-tivity of the will). Poznań: Ksiegarnia Akademicka, 1946.

9. FERRARI, G. C. Experimental psychology in Italy. Amer. J. Psychol., 1905, 16, 225–227.

10. FERRARI, G. C. Autobiography. In C. Murchison (Ed.), History of psychol-ogy in autobiography. Vol. 2. Worcester: Clark Univer. Press, 1932. Pp. 63–88.

11. FERNBERGER, S. W. Some European laboratories—1951. Amer. J. Psychol., 1952, 65, 619–626.

12. GEMELLI, A. Autobiography. In E. G. Boring, H. S. Langfeld, H. Werner, & R. M. Yerkes (Eds.), History of psy-chology in autobiography. Vol. 4. Worcester: Clark Univer. Press, 1952. Pp. 97–121.

13. GEMELLI, A., & BANISSONI, F. Speranze e preoccupazioni degli psicologi italiani in tema di insegnamento della psicologia nelle università italiane e nei vari tipi di scuole dell'ordine superiore. Arch. Psicol. Neurol. Psychiat., 1941, 2, 796–821.

14. IMUS, H. A. Experimental psychology in Italy. (Technical Report ONRL 63–52, Unclassified.) London: American Embassy, 1952.

15. KIESOW, F. Autobiography. In C. Murchison (Ed.), History of psychology in autobiography. Vol. 1. Worcester: Clark Univer. Press, 1930. Pp. 163–190.

16. LEMKAU, P. V., & DE SANCTIS, C. A survey of Italian psychiatry, 1949. Amer. J. Psychiat., 1950, 107, 401–408.

17. LORENZINI, G. L'Instituto di Psicologia Sperimentale del Pontificio Ateneo Salesiano. Salesianum, 1947, 9, 240–258.

18. MANOIL, A. La psychologie expérimen-tale en Italie. École de Milan. Paris: Alcan, 1938.

19. MARZI, A. Scuola e psicologia. Riv. Psicol., 1943, 39, 166–172.

20. MARZI, A. La psicologia in Italia dal 1939 al 1943. Riv. Psicol., 1944–45, 40–41, 193–207.

21. MESCHIERI, L. Bibliographia psicologica italiana: bollettino di informazione, 1948, No. 1. Roma: Instituto Nazio-nale di Psicologia, 1951.

22. MURCHISON, C. (Ed.) Psychological register. Vol. 3. Worcester: Clark Univer. Press, 1932. Pp. 1053–1089.

23. NAVILLE, P. Quelques aspects du dé-veloppement de la psychologie en Italie. Enfance, 1948, 1, 449–453.

24. PLOTTKE, P. Psychology in Italy. Indiv. psychol. Bull., 1946, 5, 89–91.

25. PONZO, M. In memoria di Alessandro Gatti. Riv. Psicol. norm. pat., 1938, 34, 1–2.

26. SAFFIOTTI, F. V. La evoluzione della psicologia sperimentale in Italia. Riv. Psicol., 1920, 16, 129–153.

27. WILGALIS, K. H. Die Stellung der Psy-chologie an Italienischen Universitäten und Instituten. Z. Psychol., 1938, 142, 193–199.

28. ZUNINI, G. Psicologia. (2nd Ed.) Brescia: Morcelliana, 1950. Chap. 15. La psicologia in Italia.

THE MEASUREMENT OF INDIVIDUAL DIFFERENCES IN ORIGINALITY[1]

R. C. WILSON, J. P. GUILFORD, AND P. R. CHRISTENSEN

The University of Southern California

One of the most important aspects of creative thinking is originality. This article discusses the problem of developing methods for measuring individual differences in originality. The problem arose in connection with a factor-analytic study of creative thinking conducted at the University of Southern California.[2]

In that investigation various definitions of originality were considered in the light of their implications for measurement. Three definitions and corresponding methods of measuring originality were finally adopted and applied to specially constructed tests. The methods are based upon: (a) uncommonness of responses as measured by weighting the responses of an individual according to the statistical infrequency of those responses in the group as a whole; (b) the production of remote, unusual, or unconventional associations in specially prepared association tests; and (c) cleverness of responses, as evaluated by ratings of degrees of cleverness exhibited in titles suggested for short-story plots.

These three methods permit the operations of measurement of individual differences and, while recasting the definition of originality, they preserve much of the essential meaning usually assigned to the concept. In the following sections, some of the nonmeasurable aspects of originality are pointed out and each of the three proposed methods is discussed in conjunction with a description of tests developed to utilize the method. Since the tests were included in a factor analysis along with other tests of creative thinking, the three methods are evaluated in the light of the loadings of scores from these tests on a factor which has been called originality (2).

DEFINITION OF ORIGINALITY

In developing methods for measuring individual differences in originality, the meaning to be assigned to the term *originality* and the operations for measurement must be clearly specified. The term originality has several distinct meanings. We wish to use it as the name for a psychological property, the ability to produce original ideas. What we mean by an original idea will be further specified in relation to each of the proposed methods of measuring originality.

Many writers define an original idea as a "new" idea; that is, an idea that "did not exist before." They are frequently not in agreement, however, in their interpretation of "new," since they use it with different connotations. We shall point out the inadequacy of two of these connotations for the measurement of individual differences in originality.

In one connotation, a "new" idea is an idea that "has never previously been thought of by anyone who has ever lived." In practice, of course, it would be impossible to verify whether

or not an idea meets these requirements of newness since one could never examine all the ideas of everyone who ever existed to determine whether the idea has been thought of before. This conception also presents a problem in the case of independent productions of the same idea. Two or more scientists may produce the same idea independently in different parts of the world. One of them may precede the others by a matter of months or weeks, or even hours or minutes. In trying to find creative scientists, we would probably not wish to regard the scientists who produced the idea later as unoriginal merely for having been preceded by someone unknown to them.

On the other hand, we find that "new," while meaning that which did not exist before, is sometimes interpreted, at least by implication, as including all human behavior that is not repetitive. That is, not only poetry, science, and inventions, but dreams, hallucinations, purposive behavior, and all perceptions are regarded as new. They are "new" in the sense that they are never duplicated exactly, even by the individual himself. Such a conception of "new" also fails to be fruitful, since it does not supply us with a basis for differentiating between more original and less original individuals.

For measurement purposes, we have found it useful to regard originality as a continuum. We have further assumed that everyone is original to some degree and that the amount of ability to produce original ideas characteristic of the individual may be inferred from his performance on tests. Rather than define original as "new" or "did not exist before" we have investigated three alternative definitions. We have regarded originality in turn as meaning "uncommon," "remote," and "clever." It was felt that these three definitions include significant aspects of what is commonly meant by the term original. Tests and scoring methods were developed for each of these approaches to originality.[3]

THE UNCOMMONNESS-OF-RESPONSE METHOD

Our first approach to the measurement of originality assumes a continuum of uncommonness of response. For this purpose originality is defined operationally as the ability to produce ideas that are statistically infrequent for the population of which the individual is a member. "Population" may here be regarded as any cultural group, professional group, or other aggregation of individuals having significant characteristics in common.

This definition of originality was utilized by constructing completion or open-end tests, which require the examinee to produce responses. The tests were administered to the group of individuals whose relative degrees of originality were to be determined. The responses of all the members of the group were tallied to determine their frequency of occurrence within the group. Weights were assigned to the various responses, the higher weights being given to the statistically more infrequent responses. A score was derived for each individual either by summing the weights assigned to his responses or by counting only the responses having high weights. On the basis of the score thus derived, those individuals with the highest scores were the individuals who had

given the most infrequently mentioned responses.

This procedure may be clarified by an example. The items in the Unusual Uses test are six common objects. Each object has a common use, which is stated. The examinee is asked to list six other uses for which the object or parts of the object could serve. For example, given the item "A newspaper," and its common use, "for reading," one might think of the following other uses for a newspaper: (a) to start a fire, (b) to wrap garbage, (c) to swat flies, (d) stuffing to pack boxes, (e) to line drawers or shelves, (f) to make up a kidnap note. The test is given in two separately timed parts of five minutes each. Each part gives the names of three objects and their common use with spaces for listing six other uses per object.

All the responses given by a group of 410 Air Cadets and Student Officers to each object were classified, tallied, and weighted. A system of five weights was used. A weight of 5 was assigned for the (approximately) $\frac{1}{5}$ most infrequently mentioned responses, a weight of 4 for the $\frac{1}{5}$ next most infrequently mentioned responses, and so on down to a weight of 1 for the $\frac{1}{5}$ most frequently mentioned responses. This gave a possible range of scores for each object (six responses) of 0 to 30 and a possible range of scores for the total test (six objects) of 0 to 180. The total scores actually obtained ranged from 5 to 129.

Let us consider the actual frequencies obtained for one of the objects. The 1,767 responses to the object given by the group of 410 Air Cadets and Student Officers were tabulated. One hundred and eighty-two different uses were mentioned. Eighty of these 182 uses were unique in that they were mentioned by only one member of the group. At the other extreme, one of the uses was mentioned by 173 individuals. The three most common uses mentioned, with frequencies of 173, 94, and 90, accounted for 357 responses and were assigned weights of one. The next six most common uses, with frequencies from 89 to 48, were assigned weights of two. Nine uses with frequencies from 45 to 29 received weights of three, 24 uses with frequencies from 23 to 9 received weights of four, and the 139 most uncommon uses, with frequencies from 8 to 1, received weights of five. It should be noted that there were not exactly $\frac{1}{5}$ of the total number of responses in each weight category. Because of the way in which the responses distribute themselves it is usually not possible to designate an exactly equal number of responses for each weight. It is possible, however, to achieve a close approximation.

After the weight for each response had been determined for all six objects, each examinee's paper was scored by assigning the appropriate weights to his responses and summing them. By definition, those individuals who tended to produce the most infrequently given ideas were the ones with the highest total scores and were regarded as the most original members of the group. The mean score on the Unusual Uses test was 64.0, its standard deviation was 23.5, and its alternate-forms reliability was .74.

The same procedure was applied to the Quick Responses test and the Figure Concepts test.[4] The Quick Responses test is similar to the conventional word-association test. It

[4] For more complete descriptions of tests mentioned here see (1).

consists of a list of 50 stimulus words, derived principally from the Kent-Rosanoff list and a more recent list developed by D. P. Wilson (4). The 50 words were read to the examinees at the rate of one every five seconds, the examinee being instructed to respond with the first word that came to mind. Responses of 410 individuals were tabulated for each of the 50 stimulus words. Frequencies of occurrence for each response were determined, weights were assigned, and scores derived in a manner similar to that for the Unusual Uses test. The mean score on the Quick Responses test was 99.8 with a standard deviation of 18.7. The reliability estimate was .81 as computed for odd and even items and corrected for length.

The Figure Concepts test consists of 20 simple pen-and-ink drawings of objects and individuals. Each picture is identified by a letter. The examinee's task is to find qualities or features that are suggested by two or more drawings and to list the features and the letter designations of two drawings which possess them. For example, picture A might be a sketch of a child wearing a hat, picture B might be a sketch of a woman wearing a hat, picture C might be a sketch of young birds in a nest. The examinee might give such responses as "wearing a hat (a, b)"; "young (a, c)"; "family (a, b)"; etc.

All responses of all individuals were tabulated and classified according to frequency of mention. A further breakdown was made for each response mentioned in terms of the combinations of drawings used in identifying the feature. It was noted that while there were 190 possible pairs of drawings available, certain ones were rarely used, while others were used as a source of more than one feature. Weighting of responses was thus based on both the infrequency of the response itself and the infrequency of the drawing combination used as a source of that response.

How this dual classification affected an individual's score may be seen in the situation where two individuals gave the same response (feature name), but cited different combinations of drawings. If one individual's response was derived from a drawing combination that was frequently mentioned by others in connection with that feature, the weight assigned was low. The other individual's response, if derived from a drawing combination infrequently mentioned for that feature, was assigned a high weight.

As with the Unusual Uses test, weights were assigned so that an approximately equal number of all the responses given by the group received each weight. Each examinee's responses were then assigned their appropriate weights and the weights were summed to derive the individual's total score for the test. The mean score on this test was 29.9 with a standard deviation of 12.9. Since the format of this test did not permit the direct computation of a reliability estimate, the communality of the test (.41) found in the factor analysis is offered as an estimate of a lower bound of its reliability.

In the Number Associations test the examinee is given, in turn, four different numbers (digits) and for each is allowed two minutes in which to list as many synonyms, uses, and things associated with the number as he can. For example, for the number 4 he might list coach-and-four, for, fore, foursome, quartet, etc.

The associations listed by the group were tabulated and weights were assigned in a manner similar to

that described for the Unusual Uses test. In order to try out a further variation of the uncommonness method, however, the individual's total score was derived in a slightly different manner from that previously described. Instead of summing the weights for all the responses given by the individual, his total score was derived by counting the number of responses with weights of 4 and 5. The mean score for this test was 12.5 with a standard deviation of 3.6 and an alternate-forms reliability of .57.

In the approach described in this section, we have chosen to define original as meaning "uncommon." An original idea or response is one that is uncommon or statistically infrequent, and an individual's degree of originality, as inferred from his scores on the tests described, is characterized by the degree of uncommonness of his responses.[5]

THE REMOTENESS-OF-ASSOCIATION METHOD

The second approach is in terms of remoteness of association. Originality is here defined as the ability to make remote or indirect associations. To measure originality from this point of view, tests were constructed that required the examinee to make remote associations if he responded at all. Remoteness of association was imposed by the task. Three tests of this type were constructed. The degree of originality of an individual, ac-

[5] The reader may recall that an uncommonness or idiosyncrasy score has previously been used in connection with word-association tests in the assessment of abnormalities of behavior in clinical practice, particularly of the schizoid type. The fact that such a score measures an originality factor, as we shall show later, might be regarded as support for the popular idea expressed in the words of Seneca, "There is no great genius without some touch of madness."

cording to this definition, would be manifested in terms of the number of remote associations he made.

The Associations I test presents 25 pairs of words. The associative connection between the two words is not immediately apparent. The examinee's task in each item is to call up a third word that serves as a link between them. For example:

Given:
        Indian_____money
Write on the line between these words a word that associates the two.

There are several possible words that could be used such as penny, nickel, copper, and wampum, each of which is related to both Indian and money.

The examinee's score was the number of responses given to the 25 items in four minutes. The mean score for this test was 14.0 with a standard deviation of 4.9. The odd-even reliability estimate was .87, corrected for length.

The Associations II test is similar to the Associations I except that there is more emphasis on the correct response word having two different meanings in its relationship to the two stimulus words. It is also a multiple-choice test in which the examinee must indicate which one of five letters is the first letter of the correct association.

For example:

        tree   *a b g m s*   dog
Which of the five letters is the first letter of a word that is associated with both tree and dog and has a different meaning in relation to each?

The word "bark" is the correct answer. It means the external covering of a tree and it also means the noise made by a dog. It also begins with *b* which is one of the choices, so the examinee circles the letter *b*.

The examinee's score was the number of correct responses given to 25 items in 12 minutes. The mean score was 14.0 with a standard deviation of 3.9. The odd-even reliability estimate was .62, corrected for length.

The Unusual Uses test, previously described, was also regarded as a test requiring the examinee to respond with remote associations. Since the six items composing the test were common objects, each with one well-known use, which was given, the examinee was compelled to utilize remote associations in seeking six additional uses for each object. Both a statistical-infrequency score and a simple-enumeration score were derived for this test. The correlation between these two scores was .94. There is, of course, much spurious overlap of the two scores. In view of the high correlation between the two scores and the similarity of their correlations with other tests in the creative-thinking battery, the simpler score was chosen for inclusion in the factor analysis. The mean for this score on this test was 22.1 with a standard deviation of 6.7 and an alternate-forms reliability of .80.

In the approach described in this section we have chosen to define original as meaning "remote." An original idea or response is "remote" to the extent that the individual is required to bridge an unusually wide gap in making associative responses. An individual's relative originality, as inferred from his scores on these tests, is characterized by the number of remote responses given in limited time.

CLEVERNESS

According to the third approach, originality is defined as the ability to produce responses that are rated as clever by judges. This definition requires a test that calls forth responses showing variation on a continuum of cleverness. Weights are assigned to an individual's responses in proportion to their degrees of rated cleverness.

The Plot Titles test used to measure this type of originality presents two brief stories. For each story the examinee is allowed three minutes in which to write as many appropriate titles as he can. Although relevancy rather than cleverness is stressed in the instructions, an examination of the responses of the group revealed considerable variation in the ingenuity, cleverness, or striking quality of the titles suggested.

In an attempt to develop a reliable scoring procedure for evaluating cleverness, a sample of 50 individuals was selected from the total group of 410. These 50 individuals averaged approximately six responses for each plot. The approximately 300 titles for each plot were typed on separate slips of paper. Three judges, working independently, sorted the titles into six successive piles on the basis of their judgments of the relative cleverness of the titles. Weights from 0 through 5 were assigned to the titles in the successive piles, with the high weights being assigned to the more clever titles. Agreement among the judges is indicated by the interjudge correlations (of ratings) ranging from .53 to .76. Reliabilities of test scores derived from individual judges ranged from .69 to .77. These reliabilities were computed from the two cleverness scores, one from each story, for each of the 50 individuals. The reliability computed from the composite ratings of the three judges (.76) was not higher than that for the best individual judge. Since the most reliable judge was also the one who agreed best with the other two

judges, it was decided to have this one judge do the scoring of the test for all examinees, with one of the other judges serving as a check scorer.

In an effort to simplify scoring, a study was made of total scores derived from the weights 0 through 5. That is, each test paper was scored by the number of responses at each of the cleverness levels of 0, 1, 2, 3, 4, and 5. Intercorrelations among the six scores were computed for the sample of 50 individuals. It was found that scores based on weights 0 and 1 intercorrelated well, and scores based on weights 2, 3, 4, and 5 intercorrelated well. A combination of scores based on weights 0 and 1 had a low correlation with a combination of scores based on weights 2, 3, 4, and 5. It was decided to reduce the scale to two intervals, clever and nonclever. That is, responses receiving weights 0 and 1 would be called nonclever. This greatly reduced the fineness of discrimination required of the scorer. Utilizing the titles already rated as a standard, the remainder of the tests were scored on this simple dichotomy. Two scores were recorded for each individual: the number of clever titles and the number of nonclever titles. It was decided to include both scores in the computation of the intercorrelation matrix and to determine, prior to the factor analysis, whether the cleverness and noncleverness scores were sufficiently independent to warrant including both of them in the factor analysis. The correlation between the two scores was $-.031$ and their patterns of intercorrelations with other tests in the battery were quite different; consequently, both scores were included in the factor analysis. The cleverness score (based on weights 2 to 5) emerged with a loading of .55 on the orig-

inality factor. The noncleverness score (weights 0 to 1) had a loading of $-.05$ on this factor and had its highest loading (.59) on a factor identified as ideational fluency. The cleverness score had a loading of .07 on the ideational-fluency factor.

In the approach described in this section, we have chosen to define original as meaning "clever." An original idea or response is one that is rated as clever by judges. An individual's degree of originality, as inferred from this kind of test score, would be characterized by the number of clever responses given in limited time.

DISCUSSION

The seven test scores representing the three scoring methods described were included with 46 other test scores in a battery designed to explore the domain of creative thinking. The test battery was administered to 410 Air Cadets and Student Officers. The scores were intercorrelated and 16 factors were extracted. Orthogonal rotations resulted in 14 readily identifiable factors, a doublet, and a residual. Five of the seven originality test scores emerged with loadings regarded as significant (.30 and above)[6] on one of the factors obtained. Following is a list of the tests, their scoring principles, and their loadings on the factor.

Plot Titles (cleverness)	.55
Quick Responses (uncommonness)	.49
Figure Concepts (uncommonness)	.32
Unusual Uses (remoteness)	.31
Associations I (remoteness)	.30
Number Associations (uncommonness)	.25
Associations II (remoteness)	.09

[6] The common practice among factor analysts is to regard only loadings greater than .25 or greater than .30 as significant. That is, since no means are available for determining

We have tentatively named this factor originality.[7] Another test from the creative thinking battery which should be discussed in relation to this factor is the Consequences test. This test requires the examinee to list the consequences of certain unexpected events such as the sudden abolition of all national and local laws. Two scores were derived from this test on the basis of the degree of remoteness of ideas indicated by the individual's responses. The number of remote consequences was counted for one score and the number of immediate or direct consequences for the other. It was hypothesized that the remoteness of ideas represented by the remote-consequences score might refer to something different from the remoteness of ideas required by the originality tests already mentioned. A separate factor of penetration or the ability to see remote consequences in space, in time, or in a causal chain of circumstances was therefore hypothesized. No such factor emerged in the factor analysis. The remote-consequences score of the Consequences test came out with its highest loading (.42) on the originality factor. Evidently, the remoteness of ideas represented by this test score is not different from the remoteness of ideas required by the originality tests hypothesized for originality. This finding lends additional

support to the generality of the obtained originality factor.

Inasmuch as test scores representing all three methods of measuring originality have significant loadings on this factor, we may have some confidence in its generality. Had test scores of only one method emerged on the factor, we might wonder whether the factor were specific to the particular kind of scoring method.

It should be mentioned that this factor has some appearance of bipolarity since there were a few small negative loadings of other test scores in the battery on this factor. Those test scores with negative loadings are of the kind whose "right" responses are keyed on an arbitrary, conventional basis by the test constructor. The examinee who engages in an unusual line of thought is likely to be penalized for his originality in such tests. In this connection, the essentially zero loading for originality in Associations II (as contrasted with the significant loading in Associations I) is worth mentioning. In this test, too, one "correct" answer is given credit. It may be that the original examinees think of other appropriate responses whose initial letters appear among the alternatives, and for which they receive no credit.

The fact that five of our tests designed to measure originality have in common a single factor is regarded as evidence for the potential fruitfulness of the scoring methods described for the measurement of individual differences in originality. Further work is necessary in refining the tests and in validating them against objective criteria of originality. It is felt that considerable progress has been made toward the development of objectively scored tests of origi-

---

the standard error of a factor loading, it has arbitrarily been decided to interpret the nature of factors on the basis of those loadings greater than the indicated cutoffs. The more conservative one (.30) is used by the present investigators.

[7] Hargreaves (3), in a study of imagination tests using Spearman's method of tetrad differences, found an originality group factor based on tests scored for uncommonness in a fashion similar to that described earlier in this report.

nality, with promise of satisfactory reliability.

As to the relative merits of the three approaches suggested, the uncommonness and cleverness methods have the greatest amount of the originality-factor variance but are the least economical in time and energy required to determine the scores.

In an exploratory study such as this one, expenditure of time and energy in scoring by the less economical methods may be justified in terms of the insights to be gained.

In later studies, however, it is desirable to use more economical procedures. The remoteness principle is a more economical procedure, but does not yield factor loadings as high as the less economical cleverness and uncommonness procedures. The next steps will be to revise the remoteness tests in an attempt to increase their originality variance and to seek methods of simplifying further the cleverness and uncommonness scoring procedures without decreasing their originality variance.

REFERENCES

1. GUILFORD, J. P., WILSON, R. C., CHRISTENSEN, P. R., & LEWIS, D. J. A factor-analytic study of creative thinking, I. Hypotheses and description of tests. *Reports from the Psychological Laboratory*, No. 4. Los Angeles: Univer. of Southern California, 1951.
2. GUILFORD, J. P., WILSON, R. C., & CHRISTENSEN, P. R. A factor-analytic study of creative thinking, II. Administration of tests and analysis of results. *Reports from the Psychological Labora-tory*, No. 8. Los Angeles: Univer. of Southern California, 1952.
3. HARGREAVES, H. L. The "faculty" of imagination. *Brit. J. psychol. Monogr. Suppl.*, 1927, **3**, No. 10.
4. WILSON, D. P. An extension and evaluation of association word lists. Unpublished doctor's dissertation, Univer. of Southern California, 1942.

CORRECTING THE KUDER-RICHARDSON RELIABILITY FOR DISPERSION OF ITEM DIFFICULTIES

PAUL HORST

University of Washington

Loevinger (5) has insisted, quite rightly, that the Kuder-Richardson (4) reliability formulas are actually estimates of item homogeneity as well as of test reliability. She proceeds, then, to point out that K-R 20[1] has unity as an upper limit only when the items in the test are all of equal difficulty, and regards as a serious defect of formula K-R 20 the fact that its upper limit is a function of the dispersion of item difficulties. She argues that a test of perfectly homogeneous items is justified only if there is a range of difficulty in the items. Otherwise, a single item will give the same discrimination as any number of equally difficult perfectly homogeneous items. Therefore, she contends, what is needed is a coefficient of homogeneity which has unity as its maximum value, irrespective of the dispersion of item difficulties. Accordingly, she develops a formula which has this property. It is given in somewhat different notation by

$$H_t = \frac{\sigma_t^2 - \Sigma pq}{\sigma_m^2 - \Sigma pq} \quad [1]$$

where H_t is the coefficient of homogeneity of the test; σ_t^2 the test variance; σ_m^2 is the maximum variance possible for a test which has the same distribution of item difficulties as the test under consideration; Σpq is the sum of the item variances in the test. It is clear, therefore, that as σ_t^2 ap-

proaches σ_m^2, H_t approaches unity. It has been shown by Carroll (1) in somewhat different notation that the maximum possible value of the variance of a work limit test, with score being number of items correct, is given by

$$\sigma_m^2 = 2\Sigma i p_i - M_t(1 + M_t) \quad [2]$$

where M_t is the test mean and $\Sigma i p_i$ means that the p's are ranked in descending order of magnitude, each p is multiplied by its rank order, and the sum of the products is taken.

Suppose now we attempt to relate equation [1] to one of the more familiar reliability functions. Kuder and Richardson have shown (4) that the reliability of a test is given by

$$r_{tt} = \frac{\sigma_t^2 - \Sigma pq + \Sigma r_{ii} pq}{\sigma_t^2} \quad [3]$$

where r_{tt} is the reliability of the test; σ_t^2 is the variance of the test; p is the difficulty of an item; $q = 1 - p$; r_{ii} is the reliability of item i.

The problem in using [3] is to get plausible estimates of the item reliabilities. Kuder and Richardson in deriving their well-known formula 20 assumed, in effect, that all item reliabilities are equal. Although they proceeded somewhat differently, we may, from this assumption, rewrite [3] as

$$r_{tt} = \frac{\sigma_t^2 - (1 - r_{ii})\Sigma pq}{\sigma_t^2}. \quad [4]$$

If now we let r_{ii} be the ratio of the average item covariance to the aver-

---

[1] This formula is

$$r_{tt} = \frac{n}{n-1}\left(\frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2}\right).$$

age item variance, we have shown elsewhere (2) that

$$r_{ii} = \frac{\sigma_t^2 - \Sigma pq}{(n-1)\Sigma pq}.$$ [5]

Except for a slight difference in notation, equation [5] gives the same estimate of item reliability which Kuder and Richardson derived by somewhat different methods in their formula 18. Substituting [5] in [4], we get

$$r_{tt} = \frac{\sigma_t^2 - \left[1 - \left(\frac{\sigma_t^2 - \Sigma pq}{(n-1)\Sigma pq}\right)\right]\Sigma pq}{\sigma_t^2}$$ [6]

which reduces to

$$r_{tt} = \frac{n}{n-1}\left(\frac{\sigma_t^2 - \Sigma pq}{\sigma_t^2}\right).$$ [7]

Equation [7] is, of course, the well-known Kuder-Richardson formula 20 for estimating test reliability.

Let us now examine equation [5] in the light of Loevinger's criticism. For a set of items of specified difficulties, r_{ii} will be a maximum only when σ_t^2, or the variance, is as large as possible. This maximum value, σ_t^2, is given by equation [2]. First we shall write [5] in the form

$$r_{ii} = \frac{\sigma_t^2 - \Sigma pq}{n\Sigma pq - \Sigma pq}.$$ [8]

We shall prove that [8] cannot be unity unless all of the items are of equal difficulty. It is well known that the variance of a test can be expressed as a function of the item variances and the interitem correlations, if the score is number correct and the test is not speeded. This function is

where $S_i = \sqrt{p_i q_i}$ and the r_{ij} are phi coefficients. If we prove that

$$\sigma_t^2 < n\Sigma pq$$ [10]

for items of unequal difficulty, we shall have proved that r_{ii} is less than unity, since the last terms of both numerator and denominator of [8] are equal. Since $S_i = \sqrt{p_i q_i}$ we must prove that

$$\sigma_t^2 < n\Sigma s_i^2.$$ [11]

It can readily be shown that a phi coefficient must be less than unity unless both items are equally difficult. In no case can they exceed unity. If we prove that, even though all r's in [9] are unity, the inequality in [11] must still hold for unequal S's, then we will also have proved it for the case of some r's less than unity. Assuming all r's in [9] unity, we write

$$\sigma_t^2 = (\Sigma s)^2.$$ [12]

Substituting [12] in [11],

$$(\Sigma s)^2 < n\Sigma s^2.$$ [13]

We indicate the variance of the s's by σ_s^2 and, from the standard formula, write

$$\sigma_s^2 = \frac{\Sigma s^2}{n} - \left(\frac{\Sigma s}{n}\right)^2$$

or

$$(\Sigma s)^2 = n\Sigma s^2 - n^2\sigma_s^2.$$ [14]

Substituting [14] in [13],

$$n\Sigma s^2 - n^2 - n^2\sigma_s^2 < n\Sigma s^2$$ [15]

or

$$n^2\sigma_s^2 > 0.$$ [16]

Obviously, [16] holds unless the variance of the item sigmas is zero. But the variance of the item sigmas cannot be zero unless all the item difficulties are equal,

$$\sigma^2 = \begin{pmatrix} s_1^2 & + s_1 s_2 r_{12} + \cdots + s_1 s_n r_{1n} \\ + s_1 s_2 r_{12} + & s_2^2 & + \cdots + s_2 s_n r_{2n} \\ \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\ + s_n s_n r_{1n} + s_2 s_n r_{2n} + \cdots + s_n^2 \end{pmatrix}$$ [9]

or unless some of them are equal to a constant p and all the others are equal to a constant $1-p$. But in the latter case not all of the phi coefficients could be unity. So the only case in which σ_t^2 could be as large as $n\,\Sigma pq$ is when the item variances are all equal and the phi coefficients are all unity, and this can be true only when all items are of equal difficulty.

Loevinger (6) and Johnson (3) have proposed that, in order to make a phi coefficient independent of the disparity of its item difficulties, we should divide it by the maximum phi which is possible for the two obtained difficulties. Generalizing from this rationale, we shall apply the procedure to r_{ii} in [5].

$$r_m = \frac{\sigma_m^2 - \Sigma pq}{(n-1)\Sigma pq} \qquad [17]$$

where σ_m^2 is the maximum variance of a test with specified item difficulties. We then let ρ_{ii} be the adjusted value of r_{ii} and write

$$\rho_{ii} = \frac{r_{ii}}{r_m} \qquad [18]$$

or substituting [5] and [17] in [18],

$$\rho_{ii} = \frac{\sigma_t^2 - \Sigma pq}{\sigma_m^2 - \Sigma pq}. \qquad [19]$$

But we see that equation [19] is precisely equation [1], or Loevinger's coefficient of homogeneity. This is the coefficient which she preferred to K-R 20. However, this preference does not seem entirely appropriate. As we have seen, it gives an estimate of average item intercorrelation corrected for dispersion of item difficulties. Loevinger's proposal lacks one essential step. To get a more realistic estimate of test reliability, we should substitute her formula for the estimate of average item reliability in equation [4]. Substituting equation

[19] in [4] we get

$$r_{tt} = \frac{\sigma_t^2 - \left(1 - \dfrac{\sigma_t^2 - \Sigma pq}{\sigma_m^2 - \Sigma pq}\right)\Sigma pq}{\sigma_t^2} \qquad [20]$$

which reduces to

$$r_{tt} = \left(\frac{\sigma_t^2 - \Sigma pq}{\sigma_m^2 - \Sigma pq}\right)\frac{\sigma_m^2}{\sigma_t^2}. \qquad [21]$$

Equation [21] would seem to provide a more realistic estimate of the reliability coefficient than K-R 20. It yields a higher estimate if the items are of unequal difficulty and will have unity as an upper limit irrespective of the dispersion of item difficulties. Because of the factor σ_m^2/σ_t^2, it yields a higher value than Loevinger's homogeneity coefficient given by [1].

The chief disadvantage of using [21] rather than K-R 20 is, of course, the added time required to compute σ_m^2 as given by [2]. However, the major part of the labor involved in using K-R 20 is getting the item counts for the pq values. Once the p's are obtained, it is not a great deal more work to get Σip required in equation [2] to get σ_m^2.

If we have a wide distribution of item difficulties, the difference between K-R 20 and our equation [21] may actually be quite large. Suppose, for example, we have for seven items in ascending order of difficulty, the following values: .80, .70, .60, .50, .40, .30, and .20. Suppose the variance of the test is 5.25. We have, then, the following values:

$$n = 7$$

$$M_t = \Sigma p = 3.50$$

$$\sigma_t^2 = 5.25$$

$$\Sigma pq = 1.470$$

$$\Sigma ip = 11.20$$

For the K-R 20 value we have

$$r_{tt} = \frac{7}{6}\left(\frac{5.25 - 1.47}{5.25}\right) = .84 \,.$$

For the adjusted value given by our equation [21] we must first compute the maximum variance given by equation [2]. This is

$$\sigma_m{}^2 = 2 \times 11.20 - (3.50)\,(4.50) = 6.65 \,.$$

Substituting the required values in [21], we have

$$r_{tt} = \left(\frac{5.25 - 1.47}{6.65 - 1.47}\right)\frac{6.65}{5.25} = .92 \,.$$

In this particular case, therefore, with a rectangular distribution of item difficulties ranging from .80 down to .20, our estimate of reliability is almost 10 per cent higher than that given by K-R 20. It should be emphasized that, even with the correction we suggest, we are compensating only for the attenuation introduced by the dispersion of item difficulties. We still have to remember that the basic assumption in all these formulas is that all items measure the same function, and that the failure of maximum item intercorrelation is due only to the unreliability of the items. Thus, even our formula [21] should be regarded as only a lower bound to the Kuder-Richardson type of reliability even though this lower bound may be somewhat higher than that given by K-R 20.

It should be emphasized that the type of reliability which we refer to is consistency of behavior within a very limited time interval, that is, the time interval during which the items in the test are being responded to. This is, of course, also the type of reliability implied by the Kuder-Richardson formulas. However, it should also be observed that this "short term" reliability is comparable to the correlation between two tests taken within a short time interval, e.g., at the same sitting or on the same day.

REFERENCES

1. Carroll, J. B. The effect of difficulty and chance success on correlation between items or between tests. *Psychometrika*, 1945, 10, 1–19.
2. Horst, P. Relationships between several Kuder-Richardson reliability formulas. *J. educ. psychol. Measmt*, in press.
3. Johnson, H. M. Maximal selectivity, correctivity and correlation obtainable in 2×2 contingency-tables. *Amer. J. Psychol.*, 1945, 58, 65–68.
4. Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, 2, 151–160.
5. Loevinger, Jane. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 1947, 61, No. 4 (Whole No. 285).
6. Loevinger, Jane. The technique of homogeneous tests. *Psychol. Bull.*, 1948, 45, 507–529.

MODELS FOR TESTING THE SIGNIFICANCE OF COMBINED RESULTS[1]

LYLE V. JONES AND DONALD W. FISKE

University of Chicago

Experimenters in psychology frequently encounter the problem of assessing the significance of a set of experimental results, each of which has been tested for statistical significance. Although several recent articles (**1, 2, 13**) discuss the problem, none enunciates fully the appropriate models. Nor is the problem adequately covered in any commonly used statistical text. It is the aim of the present paper to clarify the differences between the models and to provide explicit statements of the assumptions underlying them and the situations to which they apply.

The general problem can be stated simply: an experimenter has a set of two or more experimental results. For each he has applied an appropriate test of significance. He then wishes to apply a test of significance to the entire set: he wishes to formulate and test a single hypothesis about the entire set of results. One example is the evaluation of several findings, each of which is relevant to the same general hypothesis. The findings may be viewed as repetitions of essentially the same experiment. Another example is the evaluation of a set of results from the same sample, such as the analysis of test items against a criterion or the validation of several predictors against one or more criteria. Both examples illustrate types of problems which arise frequently in

psychological research, where repetition of experiments and multivariate prediction are not uncommon.

The two applicable statistical models are outlined below.

THE MODELS

The binomial model. Application of this model provides an evaluation of the significance of a set of results, one or more of which have been found to be significant. To illustrate, an experimenter has five results, two of which are significant at his preselected significance level (typically .05 or .01). He now wishes to test the hypothesis that the two significant results out of five could be expected by chance, against the alternative hypothesis that the proportion of significant results exceeds that expected on a chance basis.

The probability of one or more significant results can be determined from the binomial distribution, the expansion of $(p+q)^N$ where p is the specified level of significance, $q = 1 - p$, and N is the total number of tests of significance. From the binomial, we find the probability of obtaining by chance n or more "successes" (results beyond the specified value of p). The value of this probability is the sum of $N-n+1$ terms of the binomial expansion:

$$\sum_{s=n}^{N} \binom{N}{s} p^s q^{N-s}.$$

Wilkinson (**13**) provides tables of this probability for $p = .05$ and $p = .01$, for values of N up to 25. For

375

larger values of N, up to 50, and for other values of p, the *Tables of the Binomial Probability Distribution* (15) can be consulted in order to escape unwieldy computations. As is suggested by Brozek and Tiede (2), the normal approximation to the binomial distribution supplies a satisfactory solution so long as the product of N and p exceeds 5. Following this rule, the approximation seriously departs from the exact binomial solution for any N (number of independent statistics or results) smaller than 100 if one is using the .05 level of significance, or smaller than 500 for the .01 level.

The fundamental assumption of the binomial model is that the several experimental results are independent, that the probability value for any one result in no way influences the value for any other result. This assumption will be examined in detail below.

The chi-square model. This model is used to test the hypothesis that the composite p value for the several findings could have occurred by chance. It utilizes the numerical p values instead of classifying them as above or below a specified critical level of significance. This model also assumes the statistical independence

The model is based on the proof that any p value can be transformed to a chi-square value with two degrees of freedom, and that the sum of independent chi squares is distributed as chi square. The transformation equation is

$$\chi^2 = -2 \log_e p$$

or

$$\chi^2 = -2\,(2.3026)\,\log_{10} p.$$

The composite χ^2 is given by the formula

$$\chi^2 = -2 \sum_{i=1}^{k} \log_e p_i$$

with $2k$ degrees of freedom, where k is the number of independent probability values to be combined. Since

$$\sum_{i=1}^{k} \log_e p_i = \log_e (p_1 p_2 p_3 \cdots p_k),$$

the formula for the composite χ^2 may also be written

$$\chi^2 = -2 \log_e (p_1 p_2 p_3 \cdots p_k).$$

The product of the k separate p values is the joint probability of the k independent findings. There is no convenient test for directly assessing the significance of a joint probability.

TABLE 1

EXAMPLE OF THE CHI-SQUARE TRANSFORMATION

p	$\log_e p$	$\log_{10} p$
.04	−3.2189	8.6021−10.0000
.05	−2.9958	8.6990−10.0000
.20	−1.6094	9.3010−10.0000
	−7.8241	26.6021−30.0000

$$\chi^2 = -2\,(-7.8241) = 15.6482$$
$$\chi^2 = -2(2.3026)\,(-3.3979)$$
$$\chi^2 = 15.6480$$
$$df = 2\,(3) = 6$$
$$p < .02$$

of the several results being combined.

Suppose we have obtained the following set of p values from three com-

parable experiments: .04, .05, and .20. The composite χ^2 most conveniently can be computed using the summation formula above or the corresponding formula for common logarithms as exemplified in Table 1.

The proof for the adoption of this composite test is derived and developed by Karl Pearson (12) and by E. S. Pearson (11). It is illustrated by Kendall (8, pp. 132–133) and by Lindquist (10, pp. 46–47). Baker (1) provides an abac for determining a combined probability from the probabilities of two separate results. Gordon, Loveland, and Cureton (5) have recently published a table of $-2 \log p$ for values of p from .001 to .999 which facilitates the application of this test.

An essential difference between the two models appears when we consider the amount of information utilized. In the application of the binomial model, use is made of a simple alternative: each result in the set either is significant or it is not. The application of the chi-square model, on the other hand, demands knowledge of the exact probability value associated with each result.

An experimenter may prefer to apply the binomial method to situations where the chi-square model is appropriate. He might wish to test a hypothesis on the basis of combined results by determining whether from among N independent tests of hypothesis, n or more significant results would be expected to appear by chance. However, this procedure would involve considerable risk of erroneous interpretation, when compared with the application of the chi-square model. For example, if he were operating at the .05 level of confidence and if the p values of three results were .09, .06, and .04, he would accept the null hypothesis,

for the probability of obtaining at least one result out of three beyond the .05 level is .14. But using the chi-square transformation, the combined p is .02. On another occasion, again working at the .05 level of confidence, three results might be associated with p values of .05, .04, and .95. Under the binomial model, the null hypothesis would be rejected, for the probability of achieving at least two significant results out of three is much smaller than .05. (Wilkinson's table shows it to be .0072.) The chi-square transformation, on the other hand, leads to a combined p slightly greater than .05. Such examples indicate the major differences between the two procedures and show that the binomial method, by asking a different question, ignores some of the information contained in the set of obtained p values.

ASSUMPTIONS AND LIMITATIONS

The most limiting assumption underlying interpretation of results from the application of either model, and the assumption most likely to be unwarranted, is that of statistical independence. Independence is a necessary condition in order that, under the null hypothesis, the N results be distributed as the binomial distribution in the first model and as chi square in the second. Independence, in this sense, means that the probability associated with each one of the N results is unrelated to the probability associated with each of the other $N-1$ results. The finding that one of the results is significant must not affect the expected probability for any of the other results, which remains at .50, under the null hypothesis.

The effect of the assumption of statistical independence of data upon applications of the models can best

be assessed if we consider two kinds of situations to which the models might be applied.

Situation A. Each of the several experimental results is obtained from a *different* sample of individuals. Typically, in this situation, the several tests of a hypothesis represent replications based upon independent samples. To the extent that the samples *are* independent, either model may be applied. For the case where the samples have been randomly selected, their independence is assured.

Situation B. Each of several experimental results is obtained from the *same* sample of individuals. A typical case is item analysis, where interest resides in the degree of relationship between each of a number of items and a single criterion. The items are administered to a group of respondents, and the correlation between the criterion variable and each item is determined. Under the binomial model, the experimenter then asks if the number of significant item-criterion correlation coefficients exceeds the number expected to be significant by chance. The chi-square model would seldom be useful in this situation; when used, it would resolve the question whether the several findings support the contention that in this battery the distribution of item-criterion correlations is different from that expected by chance.

It is in dealing with this situation that our inquiry is often not legitimate, for it is seldom that the separate results are statistically independent. In the item-analysis example, independence presumes that the interitem correlations are randomly distributed around an expected value of zero, i.e., the items must be uncorrelated. Referring once again to the paper by Brozek and Tiede, we are compelled to doubt the legitimacy of using the binomial model (or, indeed, either model) for the item-analysis problem which they illustrate. No evidence is provided which would indicate that the 228 questionnaire items analyzed in their report are independent. It seems unlikely that the thousands of interitem relationships were evaluated. Since the items had been selected to focus "on a variety of personality aspects described in the literature as being characteristic of individuals who are prone to develop high blood pressure" (2, p. 339), nonzero interitem correlations surely would be expected. Not only is it a serious methodological error to apply the binomial model in this example, but also the actual conclusion is likely to be wrong. It is impossible to accept the authors' conclusion that there is "a negligible probability" of obtaining 24 or more statistics, significant beyond the .05 level, in a series of 228 item statistics. With nonzero interitem correlations, the binomial model, as applied by Brozek and Tiede, is an invalid test.

Wilkinson talks of the experimenter who tests all possible relationships in his data, possibly correlating every variable with every other one. While he apparently condones the use of the binomial method and his tables for this situation, such an application also violates the assumption. The effect of the violation, however, is considerably less extreme than in the item-analysis example. Consider an array of $n(n-1)/2$ intercorrelation coefficients among n variables. Under the conditions of the null hypothesis, these may be considered randomly selected from a sampling distribution with a population correlation of zero. The lack of independence of sample statistics may be seen when two or more particular

ample r's depart from zero: e.g., if r_{ai} and r_{aj} each are positive, the expected value of r_{ij} is no longer zero, but becomes a small positive number. If the null hypothesis obtains, the effects of dependence are slight. The binomial test, if interpreted with caution, may serve as an approximate index of independence among the n variables.

This contention is supported by the results of three sampling experiments conducted by one of the writers. Scores on each of 20 artificial variables were drawn from tables of normalized random numbers (3, pp. 295–304 and 350–359). The number of scores per variable was 50, 100, and 200, respectively, for the three experiments. In each study IBM methods were used for computation of the intervariable correlation matrix. In every case the distribution of the 190 sample r's was compared with the distribution expected under the hypothesis that these r's had been independently and randomly selected from a sampling distribution with a population correlation of zero. Several tests of this hypothesis yield no evidence against it; e.g., the proportion of r's exceeding each of several levels of significance did not significantly depart from that expected by chance. The proportion of positive r's, as contrasted with the number of negative values, suggests no extrachance effects. In short, results from none of the three studies suggest that the nonindependence of obtained correlation coefficients was a factor which noticeably disturbed the sampling distribution.

In general, when combining results by either the binomial model or the chi-square transformation, the assumption of statistical independence is one which cannot be treated lightly. Some assumptions, such as that of

normally distributed variates underlying Student's t and Fisher's F, have been shown to be of little practical importance; those tests are relatively insensitive to moderate departures from normality. In contrast, if statistical dependence is present when results are being combined, the findings from either the binomial test or the chi-square transformation are likely to be affected.

Johnson (7, pp. 170–172) illustrates the chi-square transformation by combining two tests of significance (chi square and rho) applied to the same set of data. This example appears questionable because a large correlation coefficient would usually be associated with a large chi square, although the reverse would not necessarily follow.

While the assumption of independence provides the most serious limitation for testing hypotheses concerning combined results, there are additional restrictions which should be noted. For example, while the models may be applied to any complete set of independent results, they obviously are of little value unless the separate tests of significance can be classified together on some rational basis. The application of either model to several experiments with no common aspect would be meaningless. The common problem need not be related to the content of the experimental issue; one might wish to test a hypothesis that a source of bias toward significant results resides in a given method, a particular experimental condition, or a specific laboratory. In the typical instance, however, it is nonsense to combine results bearing upon unrelated experimental problems.

It should be obvious that when applying either model all obtained results must be included in the set

analyzed. The experimenter cannot legitimately select several "promising" or "nearly significant" findings with the hope that together their combined probability will be significant.

We must recognize, of course, that the rejection of the null hypothesis for a *set* of results does not imply rejection of a null hypothesis for each result individually. Thus, if we apply the binomial model and reject the hypothesis that four results, with $p \leq .05$, would occur by chance from 25 independent studies, we may not conclude that each of the four is attributable to extra-chance factors. In combining tests of significance, we are testing a hypothesis about the set as a whole, not about the individual members of the set.

Psychologists have sometimes asked: if I have performed two independent experiments and have obtained values at the .10 and .40 levels, why should I not compute the joint probability of obtaining this pair of results by multiplying the two probabilities ($.10 \times .40 = .04$)? The difficulty resides in the evaluation of the joint probability. In the interpretation of an individual probability value, p, we make use of the knowledge that p has a rectangular sampling distribution, where every value between zero and one has an equal chance of appearing. Such is not the case for products of independent probabilities; it is the χ^2 transformation which allows assessment of a sampling distribution associated with joint probability.

The error arising from the interpretation of a joint probability as a level of significance can be illustrated intuitively. Consider four independent replications of an experiment, yielding p values of .90, .60, .30, and .10, under tests of a certain hypothe-

sis. Clearly these results provide little evidence for rejection of the hypothesis. (The χ^2 transformation, appropriate to this situation, yields a composite p greater than .40.) Yet, by finding the product of these four values one obtains $.90 \times .60 \times .30 \times .10 = .0162$, which, if erroneously interpreted as a level of significance associated with acceptance of the hypothesis, leads to a completely faulty conclusion.

In more general terms, a statistical test of significance involves the following steps: the selection in advance of an appropriate statistical model for which assumptions are tenable and the sampling distribution of the test statistic is known, the comparison of obtained data (or the statistic derived therefrom) with the model, and the resulting decision that the model does or does not fit the data, i.e., the acceptance or rejection of the statistical hypothesis. It is to be emphasized that the experimenter should specify his statistical model before he looks at his data.

THE PROBLEM OF NONINDEPENDENT DATA

Let us consider the problem of evaluating the significance of a set of experimental results when these results have been derived in such a manner as to violate the assumption of independence. When the correlations among the results depart from zero, but are unknown, there is no way in which p values associated with the individual results may be combined meaningfully. In this case we may, however, achieve an over-all test of significance by treating the set of results as a whole; i.e., we can determine one level of significance for the combined data instead of obtaining p values for each of the separate results and then combining them.

The suggested approach to this problem is that provided by Hotelling's generalized Student test (6). The most useful form of this test for psychological research probably is that which provides comparison of the sets of means of two samples on each of k variables. For example, two samples might be selected on the basis of a criterion variable, a sample of N_1 individuals high on the criterion and a sample of N_2 individuals low on the criterion. We have k variates, with unknown standard deviations and intercorrelations, each of which is measured for the $N_1 + N_2$ individuals. Hotelling's statistic, T^2, provides a decision concerning whether the two samples are discriminated significantly on the basis of k mean differences. Assuming the k variates to be distributed as the multivariate normal distribution, then the statistic

$$F = \frac{N_1 + N_2 - k - 1}{k(N_1 + N_2 - 2)} T^2$$

is distributed as the familiar F distribution, with k and $N_1 + N_2 - k - 1$ degrees of freedom. The significance of T^2 may be determined readily from tabulated values of F. While the procedures for finding the value of T^2 are discussed in several texts (e.g., 8, 14), perhaps the most readable account remains the original source (6).

Let us consider the case where k sets of measurements can be considered replications of the same variate. When systematic differences among replications can be ignored, the problem of combining separate findings disappears: the k scores for each individual can be combined and the resulting composite scores can be analyzed by the conventional t or F test. When interest also resides in replication effects, one may utilize analysis of variance to test differences not only among groups, but also among replications (9; 4, pp. 288–297). In addition to the assumption that the k measurements are replications of the same variate, these approaches are further restricted by assumptions of normality, homogeneity of error variance, and independence of variance estimates. The striking advantage of the analysis of variance design, when its use is appropriate, is the relative ease of application when compared with the computational complexity of Hotelling's multivariate procedure.

SUMMARY

This paper presents models of two designs by which an experimenter may determine the probability of obtaining a particular set of results on two or more tests of significance. The appropriate assumptions and statistical techniques are discussed. Substitute procedures are suggested for the case when the assumption of statistical independence of the several results is untenable.

REFERENCES

1. BAKER, P. C. Combining tests of significance in cross-validation. *Educ. psychol. Measmt*, 1952, 12, 300–306.

2. BROZEK, J., & TIEDE, K. Reliable and questionable significance in a series of statistical tests. *Psychol. Bull.*, 1952, 49, 339–341.

3. DIXON, W. J., & MASSEY, F. J., JR. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.

4. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

5. GORDON, M. H., LOVELAND, E. H., & CURETON, E. E. An extended table of chi-square for two degrees of freedom,

for use in combining probabilities from independent samples. *Psychometrika*, 1952, 17, 311–316.

6. HOTELLING, H. The generalization of Student's ratio. *Ann. math. Statist.*, 1931, 2, 360–378.

7. JOHNSON, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.

8. KENDALL, M. G. *The advanced theory of statistics.* Vol. II. London: Charles Griffin, 1948.

9. KOGAN, L. S. Analysis of variance—repeated measurements. *Psychol. Bull.*, 1948, 45, 131–143.

10. LINDQUIST, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.

11. PEARSON, E. S. The probability integral transformation for testing goodness of fit and combining independent tests of significance. *Biometrika*, 1938, 30, 134–148.

12. PEARSON, K. On the method of determining whether a sample of size *n* supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, 1933, 25, 379–410.

13. WILKINSON, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, 48, 156–158.

14. WILKS, S. S. *Mathematical statistics.* Princeton: Princeton Univer. Press, 1943.

15. U. S. DEPARTMENT OF COMMERCE, NATIONAL BUREAU OF STANDARDS. *Tables of the binomial probability distribution.* Applied mathematics series No. 6. Washington: U. S. Govt. Printing Off., 1949.

HISTORICAL NOTE ON THE RATING SCALE

DOUGLAS G. ELLSON AND ELIZABETH COX ELLSON

Indiana University

Sir Francis Galton (1822–1911) is generally given credit for the invention of the rating scale or at least for its introduction as a psychological measuring device. Garrett writes, "He [Francis Galton] established one of the first laboratories (in 1884) wherein mental and physical tests— mostly of the sensory-motor sort— could be taken for a small fee; . . . he introduced the rating scale and questionnaire methods later so widely used" (2, p. 58).

Garrett and Schneck say "Historically, the rating scale goes back to Francis Galton, who in 1883 published a scale for rating the clearness of one's mental imagery" (3, p. 103). Guilford also agrees with these statements. In *Psychometric Methods* he writes, "There seems to be little doubt that the first rating scale employed in a psychological problem was that of Galton, used in the evaluation of the vividness of images" (4, p. 264).

We recently found clear evidence that psychological rating scales more highly developed than Galton's were in use when Galton was a child. On a visit to the museum at New Harmony, Indiana, we found a fire-warped bronze or brass plate, accompanied in the display case by a card saying that this was one of the most interesting objects seen by the Duke of Saxe-Weimar on his visit to the New Harmony Colony. As the following material shows, this plate was actually a set of psychological rating scales which was apparently used for a kind of personality evaluation in 1826 and possibly sooner.

The Duke describes this plate in some detail in his journal, which was published in 1828, *Travels Through North America During the Years 1825– 26* (1). In the part of this account describing his visit to the New Harmony Colony, under the date of April 18, 1826, he writes:

Mr. Owen showed me two interesting objects of his invention; one of these consisted of cubes of different sizes, representing the different classes of the British population in the year 1811. . . . The other was a plate, according to which, as Mr. Owen asserted, each child could be shown his capabilities, and upon which after a mature self-examination, he can himself discover what progress he has made. The plate has this superscription: scale of human faculties and qualities at birth. It has ten scales with the following titles: from left to right, self-attachment; affections; judgment; imagination; memory; reflection; perception; excitability; courage; strength. Each scale is divided into one hundred parts, which are marked from five to five. A slide that can be moved up or down shows the measure of the qualities therein specified each one possesses or believes himself to possess (1, p. 121).

The "Mr. Owen" referred to is Robert Owen, who founded the New Harmony Colony in 1825 and left it in charge of his son, Robert Dale Owen, about a year later. Robert Owen had considerable interest in what would now be known as progressive education and child psychology.

Except for a few details, the description of the plate is quite accurate. Its dimensions are approximately $7 \times 12 \times \frac{1}{4}$ inches. Down the left-hand seven-inch side (which the Duke describes as the top) are listed the traits he mentions, except for a

few (strength, courage, self-attachment) in the upper left corner which are illegible because of fire damage. To the right of each of the ten trait names is a brass strip engraved with 100 scale markings numbered by tens. Before the damage, each strip could be slid outward independently, past an engraved index near the right-hand margin of the plate.

The notion of a rating scale is not the only one anticipated by the plate. If the slides corresponding to the several traits were simultaneously placed in positions representing the numerical ratings, they would produce what might now be called a "personality profile." Whether this idea occurred to the inventors or users of the device, we do not know. So far, with the exception of the quotation above, we have found no mention of the plate or of the rating-scale notion in the works of Robert Owen or others who had contact with him.

REFERENCES

1. BERNHARDT, K. *Travels through North America during the years 1625–28.* Philadelphia: Carey Lea and Carey, 1828.
2. GARRETT, H. E. *Great experiments in psychology.* New York: Century, 1930.
3. GARRETT, H. E., & SCHNECK, M. R. *Psychological tests, methods, and results.* New York: Harper, 1933.
4. GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.

A BRIEF NOTE ON ONE-TAILED TESTS

C. J. BURKE
Indiana University

Concurrent with the recent discussions of one-tailed and two-tailed tests by Hick (4), Jones (5), and Marks (8) there has been a disturbing increase in the use of one-tailed tests in student experimental reports as well as in published and not-yet-published manuscripts. While the popularity of one-tailed tests is undoubtedly attributable in part to the overwillingness of psychologists as a group to make use of the statistical recommendations they have most recently read, there seems to be a certain residual of bad logic, so far as both statistics and psychology are concerned, which merits examination. The writer takes the position already taken by Hick (4) in all important essentials but the argument to be presented differs, at least in emphasis, from that of Hick. It should be noted that some tests, χ^2 and F for example, are naturally single-ended. Nothing said here should be construed so as to apply to them.

Both Jones (5) and Marks (8) seem to the writer to confuse somewhat two quite different notions—that an experimental hypothesis is often directional and that an experimenter may be willing to accept a deviation of any size in the unexpected direction as consonant with the null hypothesis. We shall consider two quotations from Jones.

The model above, the test of the null hypothesis against two-sided alternatives, is the one used most often by investigators in psychology. Yet in many cases . . . it is not the test most appropriate for their experimental problems. More often than not, in psychological research, our hypotheses have a

directional character. . . . theoretical considerations allow the postulation of the direction of experimental effects. The appropriate experimental test is one which takes this into account, a test of the null hypothesis against a one-sided alternative (5, p. 44).

It is a fact that many hypotheses in psychological research, experimentally conceived, are directional for the investigator conducting the experiment, but it does not follow from this that one-sided tests should be used in experimental reports.

To amplify these considerations we point out that there are, in many experiments, two statistical decisions to be made and two different levels of confidence may be involved. The first is the decision made by the individual experimenter who frequently plans one experiment from his evaluation of a previous one. We concede that here a one-tailed test is often proper. The second is the decision which determines the place of his findings in the literature of psychology. Here the one-tailed test seems inadmissible. It is the second type of decision with which we are concerned. Marks (8) has in essence repeated from statistical sources a discussion of the Type I and Type II errors which shows that the decisions made in any statistical interpretation depend only upon the underlying populations and the rule of procedure used. Any comparison of alternative rules of procedure must take into account errors of both types, the error of rejecting a hypothesis when it is true and the error of failing to reject it when it is false, but the underlying statistical considerations do not provide automatically a criterion for the selection of one rule over another. Such a criterion is to be sought in the number and kinds of errors the experimenter will tolerate. Roughly, an acceptable criterion

is to make the over-all number of errors as small as possible and at the same time to render large and serious errors relatively impossible. Within the class of hypotheses which are considered to be directional it is likely that a one-tailed test might yield a smaller over-all number of errors than a two-tailed test, but there is, under the single-tailed rule, no safeguard whatsoever against occasional large and serious errors when the difference is in the unexpected direction. If one is less willing to commit a large error than to commit a small one, it does not follow from the theory of testing statistical hypotheses that the experimenter's expectation of a given direction for the result necessarily makes the one-tailed test desirable.

To advance this point in our case against the use of the one-tailed test in the public report, we next take up the second quotation from Jones.

It might be noted that with this formulation of the one-tailed test there is no allowance for the possibility that the true difference . . . is negative. In the type of problem for which the one-tailed test is suited, such a negative mean difference is no more interesting than a zero difference (5, p. 45).

This statement is perfectly correct.[1] If we consider it carefully we discover its import to be that the investigator should use a one-tailed test when he is willing to accept a difference in the unexpected direction, *no matter how large*, as consonant with the hypothesis of zero difference. This is quite a different matter from using a one-tailed test whenever the direction of the difference is predicted, on some grounds or other, in advance. It is to be doubted whether experimental psychology, in its present state, can

---

[1] In his subsequent discussion, Jones spoils the force of this point by confusing hypotheses to be tested with classes of hypotheses to be guarded against as alternatives.

afford such lofty indifference toward experimental surprises.

The questions raised by the one-tailed test are to be answered finally by considering the effect of general use of this procedure on the content of psychological literature. The writer cannot agree with Hick (4) that its use makes little difference since there is no practical rule for deciding what "significance" really is. In some super-scientific world this point might be well taken, but there is evidence that in our workaday world (where we sometimes read only the concluding sections of reports) it does make a difference whether the investigator has stated that his results were significant. The controversy over the Blodgett effect is a case in point (1, 2, 6, 7, 9, 10, 11, 12).

Remembering that the problem of testing a statistical hypothesis is a statistical problem in which each individual experiment is viewed only as a member of a class of similar experiments and recalling that the properties of any statistical test are determined solely by the procedure followed and by the populations underlying the class, it is pertinent to inquire into the effects of widespread adoption of one-tailed tests upon the literature. The writer believes the following statements to be reasonable forecasts.

1. The discovery of new psychological phenomena will be hindered. Our literature abounds with instances in which the outcome of a given experiment has differed reliably and sharply from expectation. These experiments are usually of great interest—new psychological concepts arise from them. Our science is not yet so mature that these can be expected to occur infrequently. The most recent instance, known to the writer, of conflicting results from experiments thought to be highly similar was reported by Underwood (13) at the 1952 meetings of the American Psychological Association. From any careful examination of contemporary psychological literature we must conclude that nowhere in the field can we have sufficient a priori confidence in the outcome of any genuinely new experiment to justify the neglect of differences in the unexpected direction.

2. There will be an increase in barren controversy. Fruitless controversies arise from unreliable results. Conclusions at low levels of confidence tend to be unreliable, and the adoption of one-tailed tests is equivalent to a general lowering of levels of confidence. At a time of severe journal overload this is especially pernicious. There is no substitute in statistical methodology for the carefully designed and controlled experiment in which any important difference between groups will show up at a high enough level of confidence to insure a certain reliability in the conclusion.

3. Abuses will be rampant. It is no criticism of the position held on statistical grounds by Jones and Marks to point out that the considerations involved in the choice of a one-tailed test are really rather delicate. A nice instance of what can happen is seen in an experimental report by Gwinn (3). Gwinn reports two experiments which are not markedly different from each other. They turn out in opposite directions, and, by appropriate selection of the position of his "critical tail," Gwinn establishes significance and near significance (1 per cent and 8 per cent levels, approximately) for his results on the basis of one-tailed tests.

The moral can be pointed with advice. We counsel anyone who contemplates a one-tailed test to ask of himself (before the data are gathered):

"If my results are in the wrong direction and significant at the one-billionth of 1 per cent level, can I publicly defend the proposition that this is evidence of no difference?" If the answer is affirmative we shall not impugn his accuracy in choosing a one-tailed test. We may, however, question his scientific wisdom.

REFERENCES

1. BLODGETT, H. C. The effect of the introduction of a reward upon maze performance of rats. *Univer. Calif. Publ. Psychol.*, 1929, 4, 113–134.

2. BLODGETT, H. C. Reynolds' repetition of Blodgett's experiment on "latent" learning. *J. exp. Psychol.*, 1946, 36, 184–186.

3. GWINN, G. T. Resistance to extinction of learned fear drives. *J. exp. Psychol.*, 1951, 42, 6–12.

4. HICK, W. E. A note on one-tailed and two-tailed tests. *Psychol. Rev.*, 1952, 59, 316–318.

5. JONES, L. V. Tests of hypotheses: one-sided vs. two-sided alternatives. *Psychol. Bull.*, 1952, 49, 43–46.

6. KENDLER, H. H. Some comments on Thistlethwaite's perception of latent learning. *Psychol. Bull.*, 1952, 49, 47–51.

7. MALTZMAN, I. The Blodgett and Haney types of latent learning experiment: reply to Thistlethwaite. *Psychol. Bull.*, 1952, 49, 52–60.

8. MARKS, M. R. Two kinds of experiment distinguished in terms of statistical operations. *Psychol. Rev.*, 1951, 58, 179–184.

9. MEEHL, P. E., & MacCORQUODALE, K. A failure to find the Blodgett effect and some secondary observations on drive conditioning. *J. comp. physiol. Psychol.*, 1951, 44, 178–183.

10. REYNOLDS, B. A repetition of the Blodgett experiment on "latent" learning. *J. exp. Psychol.*, 1945, 35, 504–516.

11. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, 48, 97–129.

12. THISTLETHWAITE, D. Reply to Kendler and Maltzman. *Psychol. Bull.*, 1952, 49, 61–71.

13. UNDERWOOD, B. J. The learning and retention of serial nonsense lists as a function of distributed practice and intralist similarity. Paper read at Amer. Psychol. Ass., Washington, D. C., September, 1952.

A NOTE ON THE RECOGNITION AND INTERPRETATION OF COMPOSITE FACTORS

WAYNE S. ZIMMERMAN
Brandeis University

French's recently published monograph, *A Description of Aptitude and Achievement Tests in Terms of Rotated Factors* (1), fills a very great need; so great, in fact, that we should expect wide use to be made of it, not only by trained factor analysts, but by counselors, teachers, and personnel workers in business and industry. Particularly because of this anticipated wide use, I was very much disturbed to note the frequency with which test loadings, reported on a single factor, were actually earned on composite factors. Consequently, as reported, these loadings are in many cases greatly inflated. The serious distortion that results colors the monograph throughout, but is especially apparent in Part II.

In Part II, brief descriptions of factors are presented, accompanied by test names listed in the order of magnitude of their loadings. This

section will more than likely receive the greatest application, especially by those interested solely in selecting the best test of a particular factor. It is probably safe to assume that few of the relatively untrained, and perhaps not all of the trained, people will check carefully to see if the particular test concerned achieved its loading in analyses where a full complement of factors in the related subdomains was isolated.

In his introduction French discusses composite factors, but in such a way that the reader may be left with the impression that these factors occur very infrequently, or that when they do their nature is apparent and they do not therefore seriously influence the monograph presentation.

Composite factors may occur when several tests included in a correlational matrix have a similar factor pattern. If, at the same time, no other tests have a sufficiently disparate pattern on the same factors, a separation between factors cannot be achieved, even when an ample number of centroid factors are extracted.

Composite factors may be produced also when a curtailed number of centroid factors are extracted. In too many instances tests have been included for analysis which actually represent a greater number of factors than can be extracted according to any criterion applied to determine when the extractions should cease. In many other instances a too rigid criterion is employed. The experience of some analysts has shown that the use of additional centroids, extracted after the most liberal statistical criterion is satisfied, frequently brings about more readily a psychologically meaningful solution.

In this short critical note, I can do no more than call attention to one or two of the most glaring examples of the failure to note the accretion because of composite factors to certain reported test loadings.

On the Space factor it is necessary to work down the list through eleven tests before locating one with a loading that seems to be relatively free of the composite taint. The loading for the leading test is based upon a single analysis only—an analysis in which just four factors were extracted with only one in the spatial-visualization domain. AAF analyses of a nearly similar test (1, AFL-7; 2, No. 32) have demonstrated that where other related factors (i.e., Perceptual Speed, Length Estimation, Visualization) are isolated, the test variance is spread among them leaving a very much reduced weight on the Spatial factor. The position of the next three tests in the list can be questioned by reference to this same list where a combination of the three tests into a single form is ranked twenty-fourth, with considerably less than one-half of the variance. This circumstance suggests that there is probably a significant amount of specific variance represented when any of the three highly related tests are intercorrelated and analyzed as single tests. At the same time a much reduced Space loading was found for this combined form in analyses where both Visualization and another factor in the spatial-visualization area (1, Factor SO) appeared also (1, AFM-21, AFN-21). The next seven tests, which include several of the form-board type, were all analyzed in batteries in which only a single spatial-visualization factor was isolated. Where more than one related factor was revealed, the Space variance of similar tests was reduced very markedly. (See 1, ThA-21, Fr-10, M, B-14. Also AFE-15, AF6-8, AFM-23.)

Of the first ten variables in the list of Deduction tests, five fairly obviously are composites, since three

represent grades assigned in academic subjects and two are based upon comprehension of school subject matter. Two other tests actually achieved the loadings credited to them on a factor that was called Attention by the author of the analysis. One test was analyzed with two other forms of the same test, both of which produced one-half or less variance on the same factor. The remaining two test loadings were based on a single score derived on three different tests, each of which singly in other analyses demonstrated a different and complex factor pattern. In none of the analyses in which these first ten tests were analyzed was a representative group of factors in the reasoning area isolated.

It seems that there are still factor analysts who, although they accept rotation as a necessary step in the process, somehow are not aware of one implication from the pattern (often indistinct and sometimes enmeshed beyond recognition, it must be granted) that tends to form in the centroid loadings. It is well known that in dealing with intellectual variables the first centroid loadings are usually all positive and the second centroid, as well as those that follow, divide positive and negative variance equally. What is apparently overlooked is the tendency for the second centroid to split the most obvious dichotomy, the third centroid to split the next most obvious dichotomy, and so on. For example, if the battery contains both linguistic and quantitative tests, the second centroid will most likely separate these two major groups. In order to effect further separations between test variables on the linguistic end of this second axis (which, in pair-by-pair rotation, has now been extended by rotation with Centroid Axis I) it must be rotated with the next centroid that shows this new separation most clearly. Similarly, further separation of the quantitative tests must be achieved through rotation of the quantitative end of the second axis (also extended through its rotation with Centroid Axis I) with a centroid which in turn shows the most distinct separation among the quantitative tests. This separating process must be continued until all the factors are accounted for. If too few centroid factors are extracted, some of these separations cannot be made. Consequently, a test which actually contains variance on two or more factors may appear with all of that variance confined on a single factor.

It is my feeling that the failure to give composite factors the attention they merit must be considered either a serious oversight or a serious error of omission. Since the user must make his own evaluation, I hope this note will draw attention to the necessity of considering very carefully the serious effect of composite factors upon certain test loadings reported in French's monograph.

REFERENCES

1. FRENCH, J. W. *The description of aptitude and achievement tests in terms of rotated factors.* Chicago: Univer. of Chicago Press, 1951.

2. GUILFORD, J. P., FRUCHTER, B., & ZIMMERMAN, W. S. Factor analysis of the Army Air Forces Sheppard Field battery of experimental aptitude tests. *Psychometrika*, 1952, 17, 45–68.

A REJOINDER TO ZIMMERMAN'S NOTE

JOHN W. FRENCH

Educational Testing Service

Zimmerman's criticism of the treatment of composite factors in *A Description of Aptitude and Achievement Tests in Terms of Rotated Factors*[1] brings up an important point. Analyses using very few tests often yield combination factors that seem to be composed of two or more factors ordinarily found separately. The tests found to have high loadings on these combination factors are often complex. When such complex tests are analyzed in larger batteries, they often receive lower loadings on each of the separate factors.

Now, what should be done about this in comparing loadings from many analyses, large and small? The centroid of the high-loading tests for a particular factor in one analysis is probably never exactly duplicated in any other analysis. Nevertheless, it is necessary to make some cross-identifications, provided the factors are close enough to being alike. How close is that? If you insist upon extreme closeness, you identify so few factors that the result is not useful. If you are very liberal, the factor loadings cannot be compared because they are not loadings on the same thing. The combination factors mentioned by Zimmerman cause most of the trouble, but factors do not always combine equally. For example, consider the case of an analysis in which neither of the well-known factors A or B appears, but in which tests ordinarily found on A and B are now found on the same factor. As usually happens, the division is not exactly equal; the A-factor tests may be more numerous and highly loaded. Perhaps a combination factor correlating .90 with A and .40 with B can safely be identified as A, while a combination factor correlating .80 with A and .60 with B should be left unidentified. Even in the former case a test having loadings on both A and B will, unfortunately, tend to have a higher loading on the impure A factor than do the tests which best represent pure A. Thus, a compromise must be made between too few identifications, on the one hand, and too poorly representative loadings, on the other.

Zimmerman's note suggests that listing the tests according to the size of their loadings is misleading. They were listed in this manner to be entirely objective. How else should they be listed? Perhaps they should have been listed in accordance with some value derived from differentially weighting such considerations as size of loadings, number of tests in the analysis, number of analyses in which the test appears, number of subjects, estimation of closeness of factor to other factors identified by the same name, closeness of the test in psychological meaning to that of the factor, etc. To do this would have been to put more reliance on my judgment with regard to weights than I cared to put, and would have increased the effects of any possible misinterpretations of mine beyond the degree in which they already appear.

One answer, as suggested by Zimmerman, is to wave more red flags. I am now preparing a similar exposition of personality factors, where the problem of combined factors is even worse. I will have to give this problem much attention, and, because of his note, will take extra pains to warn the nonfactorists of possible pitfalls. I wish Zimmerman were right in saying that the monograph serves the masses, but the figures on sales indicate that we have a most exclusive group of readers.

---

[1] FRENCH, J. W. *A description of aptitude and achievement tests in terms of rotated factors.* Chicago: Univer. of Chicago Press, 1951.

DE GRAZIA, SEBASTIAN. *Errors of psychotherapy.* Garden City: Doubleday, 1952. Pp. 288. $3.00.

LINDNER, ROBERT. *Prescription for rebellion.* New York: Rinehart, 1952. Pp. 305. $3.50.

These two books, when read together, make an interesting contrast. After a lapse of a week or two, however, they become strangely intermingled in memory.

Each is a sort of intellectual *tour de force.* Each is the work of a literate, passionate prophet who, starting with relatively little data, evolves a logical structure which becomes a formula for saving humanity from itself. Both authors have little good to say for contemporary psychotherapy and each sets forth his convictions concerning the only way things may be set aright before it is too late. The fact that their peregrinations lead to markedly different conclusions does not dissipate the inhibition the memory of one book has on the memory of the other.

Both authors are sharply critical of contemporary psychology and psychiatry. They are critical, though for different reasons, of traditional psychoanalysis and nondirective therapy, and they heap scorn on shock therapy and lobotomy.

Both authors mistake hypotheses for conclusions. Each ends his book with generalizations which, in more modest men, might lend themselves to testable hypotheses. Hoping for data, one runs head-on into the back cover.

de Grazia's message is fairly simple. Neurosis is a moral disorder. The immoral psychotherapists, worshiping false gods, do not possess the kind of authority essential for forgiveness of moral transgression. For-

giveness is essential, however, to valid therapy—mere toleration degenerates whatever moral standards remain. The book ends with a description of the true psychotherapy which, it turns out, requires periodic confession to a religious authority who has power to recognize and forgive moral transgression. In the process of building his logical structure, de Grazia examines the status of confession in various religions and the consequences of its loss or retention. About the most convincing datum bearing on the value of confession is a quotation from a Hawthorne novel. De Grazia debunks the available data on evaluation of therapy (what *is* cure?) and he exposes the error in such concepts as unconscious motivation, psychological determinism, developmental factors in neurosis, and other sacred psychological cows. To some psychologists he will not appear to lean over backwards to be friends. In his introduction, he anticipates that some will say that he has set psychotherapy back twenty-five hundred years. This estimate may be high.

de Grazia writes with a sharp stylus. Lindner uses a half-inch camel's-hair brush and mixes more than the usual amount of oil with his paints. Lindner, too, is critical of modern psychotherapy, but for a different reason. To him psychotherapists are succeeding in making people adjusted. This is terrible. Adjustment is conformity and stereotypy and leads to Mass Man and totalitarianism. Lindner shares with his readers his discovery that the goal of true psychotherapy and, more broadly, of social therapy is rebellion. He builds his case for individuality and creativity. He decries passivity and homogeneity. He places psychologists at

the head of the list of the culprits leading mankind into bondage under the banner of adjustment. He exposes psychology's hypocrisy and criticizes its guiding fictions. In the end he reveals his prescription for the new psychotherapy. Psychotherapy must have as its goal and criterion of success "the re-entry of the individual . . . into the evolutionary stream." He does not specify the exact units to be used in measuring this variable.

Many psychologists reading either of these two books will find themselves muttering objections and writing comments in the margins. Despite this they will find much that is stimulating. Both authors express their ideas clearly and with great certainty. Their readers may be less certain but they cannot avoid having their thinking provoked.

GEORGE W. ALBEE.
Washington, D. C.

BENEDEK, THERESE: *Studies in psychosomatic medicine: psychosexual functions in women.* New York: Ronald, 1952. Pp. x+435. $10.00.

This is the second volume in the series of psychosomatic studies sponsored by the Chicago Psychoanalytic Institute. The first eleven chapters reprint the 1942 monograph published by the National Research Council. The last four chapters are psychoanalytical essays on some female psychosexual functions. Since no new data, methodology, or analyses are appended to the reprint, the evaluation of the major portion of this book is in essence a review of the 11-year-old monograph.

Two earlier studies led Benedek, and her collaborator Rubinstein, to conclude that in the sexual cycle there is a dynamic correlation between each hormonal variation and the psychodynamic manifestations of the sexual drive; that the sexual cycle could be adequately described only if the emotional cycle and the hormonal cycle were both included as integral components.

The present study is an extension of the earlier investigation. Daily psychoanalytic therapy productions and daily physiological hormone indices were collected from 15 psychologically disturbed patients. These data were combined with the data of the two earlier studies making a total of 152 female sexual cycles. Six of the patients contributed two-thirds (101) of the cycles. Although the vaginal smear was not a direct hormone assay, biophysiological studies by Rubinstein gave reasonable assurance that the smear was an adequate first approximation of the hormone level.

The daily therapy data were psychoanalytically interpreted, and classified in terms of three groups of predominant "psychodynamic tendencies." From these psychodynamic tendencies the daily hormonal characteristics were predicted: (a) Estrogen production was inferred when the patient presumably showed active psychodynamic tendencies directed towards an externalized object (e.g., heterosexual tendencies, masculine identification, active homosexual tendency, infantile sexuality, oedipus complex, masochism-sadism, and exhibitionistic tendency). (b) Progesterone hormone levels were inferred when passive receptive psychodynamic tendencies could be interpreted from the daily material (e.g., libidinous narcissism, genital receptivity, retentive wishes, pregnancy wishes, mother-conflict and mother-identification, motherliness, and passive homosexual tendency). (c) Low hormone levels were inferred when the therapy data permitted the interpretation of such psychody-

namic tendencies as destructive hostility, negative narcissism, infantile demanding dependence, depressive and eliminative tendencies.

Certain methodological unclarities suggest caution in accepting the author's claim of high predictive accuracy. Benedek does not state whether the individual days were analyzed in random order, nor how the influence of knowledge about related days in the cycle was guarded against in the evaluation of the given day. These points are particularly pertinent inasmuch as Dr. Benedek herself states that the interpretation of any psychological production requires a knowledge not only of the given sexual cycle, but also of the complete clinical material for the patient. There are no data as to the reliability of the criteria for determining psychodynamic tendencies either for one judge or for independent interpreters. It is also regrettable that no definite statement is made as to the specificity of the hormone predictions which the psychological material could be expected to yield. At times it appears from the data in Tables 17–42 that a three-point scale is used (estrogen, progesterone, low hormone level); yet at other times a more elaborate scale is used (incipient estrogen, estrogen, high estrogen; incipient progesterone, progesterone, declining progesterone, low hormone level). This creates a problem in understanding how the author arrived at her summary statements of accuracy of prediction. The criteria for partial and total discrepancy between predicted hormone level and vaginal smear hormonal index are sometimes difficult to interpret: psychological prediction *Progesterone*, vaginal smear finding *Estrogen* is listed as a partial discrepancy (p. 322); psychological prediction *Estrogen and Minimal Pro-*

gesterone, vaginal smear finding *Increasing Estrogen* is listed as no discrepancy (p. 293).

There seems little doubt, however, that Dr. Benedek was able to infer the probable hormonal status from the psychological productions of these particular subjects at far beyond chance levels. This is an important clinical demonstration of psychosomatic unity of function. But the lack of scientific modesty in limiting the broad generalizations and conclusions derived from the data casts a shadow over the positive contribution of this monograph.

The first of the four essays, Mother-Child as a Primary Unit, deals with " ... the psychodynamics of the symbiosis which exists during pregnancy, is interrupted at Birth, but remains a functioning force, directing and motivating the mental and somatic interaction between mother and child." The hormonal functions which are related to milk secretion induce an emotional attitude in the mother in which active and passive receptive tendencies predominate in a manner comparable to the progesterone phase of the feminine sexual cycle. This is said to promote the development of motherliness and to foster the development of the personality of mother and child by gratification of the dependent needs of both. The apparent regression to an oral level by the mother in her unconscious identification with the infant presumably is the psychodynamic tendency which predisposes to depressive reactions in the puerperium.

The Climacterium, topic of the second essay, is characterized as a period of intrapersonal reorganization in women. The adequacy of the earlier psychosexual adjustment from menarche through motherhood becomes

the important source of "surplus gratification." Thus, the end of the propagative function may be viewed as the release of nonsexualized energy for a new phase in which an active, adaptive life plan can be developed.

The third paper, The Functions of the Sexual Apparatus and Their Disturbances, is a potpourri of psychoanalytic clichés and a recapitulation of the essential conclusions of the monograph and of the first two essays in this series of four. The discussion of dysmenorrhea is the only new contribution. Dysmenorrhea is treated without differentiation as to whether it is primary or secondary in character. The psychoanalytic explanation of dysmenorrhea is in terms of a diminished control of the ego at the time of menstruation over psychosexual conflicts which by "returning from repression" mobilize anxiety and general nervous system reactions and predisposes the woman to overreact to the premenstrual-menstrual hormone changes.

The final essay, Some Psychophysiological Problems of Motherhood, follows the pattern of reviewing the monograph and the first essay in the series. The generalizations about "motherliness" conflicts appear to be derived from upper-middle class women who can afford psychoanalytic treatment, yet these generalizations are freely expanded to include all women. There is a high degree of autonomy in these essays, and their juxtaposition in print seems more an accident of convenience than a plan.

The republication of the 1942 monograph to make it more available to a wide audience is to be welcomed. The clinical methodology, the efforts to elaborate criteria for judging psychodynamic tendencies, and the demonstration of psychosomatic unity of functioning are of value to research

psychologists, physiologists, and physicians. The brief summaries of psychoanalytic theory introduced for the nonpsychiatric reader are excellent condensations, succinct yet thorough.

M. Erik Wright.
University of Kansas.

Maier, Norman R. F. *Principles of human relations.* New York: Wiley, 1952. Pp. ix+474. $6.00.

This text on human relations extends the trend of publications in this field. Industrial personnel now realize more than ever before the need for scientific techniques and approaches in the promotion of better employer-employee relations. Very often, a human relations program in industry is based on the experiences of the personnel man or upon half-thought-out methods provided by "pulp" magazines in the field of supervision and management. Industry has grown tired of these approaches and now wants something scientific upon which to base its programs. Maier's text fulfills this need to some extent.

The book has been aimed primarily at the industrialist. It is typically a Maier book on frustration, morale, and attitude-change techniques. Approximately 90 per cent of the book is devoted to attitude-change techniques. Little attention is placed on experimental results. It is in this area that the greatest weakness is shown. The author tends to draw too many conclusions and overgeneralizes from single case studies. A feeling develops that techniques and approaches are colored by the background and training of the author. Basic studies by Lewin, Bavelas, Coch and French, and Lippitt are cited, but are these enough to pave the way for the application of the group decision methods in industry?

Maier's chapters on role-playing and group-decision procedure are undoubtedly the best written in the field. Possible applications to industrial personnel problems are well made. The book will be well received in industry. He has many case studies to bring home his points.

One wonders why nondirective counseling is discussed in his last chapter. Unfortunately, this introduces the reader to a complicated area of human study that might have best been eliminated or included as one of the interview methods.

In summary, the book is very timely in its presentation and undoubtedly is what is now needed in industry. It will provide personnel people with some concrete basis for this field of study.

D. J. Moffie.

North Carolina State College.

Brower, Daniel, and Abt, Lawrence E. (Eds.) *Progress in clinical psychology.* Vol. I. New York: Grune & Stratton, 1952. Section 1: pp. xi+328; $5.75. Section 2: pp. xxiii+236; $5.00.

This is the first volume, in two sections, of a projected series of periodic reviews of the literature in clinical psychology. Section 1 contains Parts I, II, and III: Introduction, Diagnostic and Evaluative Procedures, and Psychotherapy. Section 2 contains Parts IV, V, VI, and VII: Developmental Processes, Applications of Clinical Psychology to Special Areas, Approaches to Clinical Psychology, and Professional Issues. The range of the material encompassed is so great as to preclude a detailed critical evaluation. One can, however, examine the basis for selection of material, the general theoretical approach, and perhaps comment briefly on some of the 42 chapters.

In the preface to the first section, the editors state that "We early reached a decision to employ value judgments in the selection of material to be covered. . . . " Aside from the subsequent statement that "We have included what we believe to be a judicious selection of the varied, complex and interlaced applications of clinical psychology in special areas of concern and interest . . . ," the reader is nowhere told what these values are. If the content of this volume is a reflection of the editors' value judgments, a psychoanalytic orientation predominates. This is probably inevitable since psychoanalysis has provided the most valuable hypotheses for psychodiagnosis and psychotherapy. However, other points of view are represented, as in the chapter by Raymond B. Cattell on P technique and the conventional discussion of intelligence testing in the child by Dale B. Harris, among others.

It is the psychoanalytic orientation apparently which leads Abt, in the opening chapter, to attribute the delay in the development of psychology relative to the other sciences to the fear that men have of discovering their own natures. In the reviewer's opinion, this is a naive view of the history of science, and represents a lack of sobriety in the application of psychoanalytic concepts, which has been abandoned even by many psychoanalysts.

As can be expected in any large collection of papers, they vary widely in quality. There are a number of excellent critical reviews of the literature. For example, the chapter on Psychosurgery by Birch is outstanding because of the conciseness of its exposition and the acuteness of its insight into the basic lack of sound supporting data. With particular reference to the increasing use of psychosur-

gery as a method of choice rather than of last resort, Birch concludes that "the evidence indicates the need to halt the ever-widening use of a radical practice that has neither a clear theoretical justification nor a sound empirical base."

Hertz's chapter on the Rorschach is a thorough and competent review of the literature. There is an absence of defensiveness with respect to experimental data which makes it particularly appealing. At the same time, she points out that the need for experimental research should not blind the clinician to the equal need for *art* and intuition in the interpretation of an individual Rorschach protocol. In sharp contrast stands the rather desperate defense of the Szondi test by Deri, who attacks all critical studies, and omits consideration of some of the most damaging ones. Her plea for tolerance of ambiguity by clinicians thus fails to strike a sympathetic chord.

Clinical and other applied psychologists should find a number of the individual papers informative, particularly those on topics with which they may deal only peripherally in their everyday practice. For this reviewer, the chapters on Measures of Aptitude, Achievement and Interest; Testing for Psychological Deficity; House-Tree-Person and Human Figure Drawing; Mental Deficiency; and Cultural Anthropology, by Gustav, Hunt, Brown, Heiser, and Sargent, respectively, were in this category. It is doubtful, however, that the general academic psychologist will find much of lasting theoretical value which cannot be obtained elsewhere. Abt's chapter on The Emergence of Clinical Psychology is interesting but sketchy. Mowrer's chapter on Neurosis and Its Treatment as Learning Phenomena is a restatement of his

well-known position on the "neurotic paradox." Else Frenkel-Brunswik's chapter on Social Psychology emphasizes the contributions of psychoanalysis to this area.

Some of the authors cite references and findings without adequate comment. The reader is often left in the midst of a maze of numerous journal references. The two chapters by Ellis on Self-Appraisal Methods and Sexual Disorders are saved from this fate, despite the inclusion of more pages of closely printed references than of loosely printed text, by the author's summaries and integration.

On the whole, there is much information of interest in these two books, most of which, however, can be found in other integrated sources, such as the *Annual Review of Psychology*. The question may be raised therefore whether these volumes add sufficiently to available sources to warrant a new series of periodic reviews of the literature, even if in a specialized area.

<div style="text-align: right">Julius Wishner.
University of Pennsylvania.</div>

Lansing, A. I. (Ed.) *Cowdry's problems of ageing.* (3rd Ed.) Baltimore: Williams and Wilkins, 1952. Pp. xxiii+1061. $15.00.

To the psychologist, the extent to which psychological data are omitted from this book is amazing. While our contributions to the understanding of ageing have not been spectacular, they are not nonexistent as would appear from this volume. No mention is made of researches on age changes in speed and accuracy of response, learning, creativity, memory, vocabulary, or mental abilities of any sort. Of 40 chapters, not one is psychological.

Some fields closely related to psychology are likewise ignored. For ex-

ample, although there are three chapters on the sexual organs, sexual behavior is barely mentioned. I am unable to find any reference to Kinsey. Likewise there is no treatment of psychopathology, although a sharp increase in mental disease in later maturity is one of the prominent features of ageing.

The orientation of the volume is clearly medical; but medicine, it seems, does not include psychiatry. The treatise is not purely biological and medical; there are four chapters by sociologists and economists, but it could scarcely have been the richness of the literature on ageing in these disciplines which prompted the inclusion of these chapters. To illustrate, the chapter by Havighurst on Roles and Status of Older People cites only three investigations. Two of these are by Havighurst and one is a University of Chicago dissertation. Such poor documentation is not, however, characteristic of the medical and biological chapters.

The heterogeneity of this book illustrates the fact that there are no gerontologists, but only men specializing in the gerontological aspects of particular fields. In this situation, it would seem to have been wiser had the decisions concerning areas to be included been made by a board of editors representing many disciplines.

WAYNE DENNIS.

Brooklyn College.

PIÉRON, HENRI, PICHOT, PIERRE, FAVERAGE, J. M., AND STOETZEL, JEAN. *Méthodologie psychotechnique.* (Book II in *Traité de psychologie appliquée.*) Paris: Presses Universitaires de France. Pp. viii +208. 1.000 fr.

This is the second volume of an ambitious treatise on applied psychology, published under the editorship of Henri Piéron and planned in seven volumes. The first volume, issued in 1949 and entitled *Differential Psychology*, dealt with the concepts of individual variability and "types," aptitudes and their factorial analysis, and the role of heredity. The present volume is concerned principally with the methods of applied psychology.

Piéron's section contains a general introduction, concerned both with tests as tools of measurement and with the evaluation of the test scores, and an inventory of principal psychometric tests. The European reader would have profited by being given a reference to sources of information on the vast amount of American work on the construction and utilization of psychological tests, other than Whipple's Manual (1910), Bronner *et al.* (1928), Garrett and Schneck (1933), and Rapaport (1945). References to individual tests and to works on test theory are more up-to-date. Piéron's chapter will acquaint the American student with developments in the field of testing in Europe, especially in France. Of particular interest and potential usefulness is a nonverbal (pictorial) group test of intelligence designed for children of school age and developed by René Gille.[1] The test was standardized on a sample of about 100,000 French boys and girls aged 6 to 12 years and was adapted for use with the African Negro children. Selected tests of sensory, perceptual, motor, and a variety of "mental" functions (such as attention, memory, and spatial orientation) are described, mostly very briefly.

[1] Cf. Heuyer, G., Piéron, H., Piéron, Mme. H., and Sauvy, A. Le niveau intellectuel des enfants d'âge scolaire (Intellectual level of children of school age). Paris: Institut National d'Etudes Démographiques, 1950. Rep. No. 13 of *Travaux et documents.*

Methods for the study of personality are presented by Pichot. The methods are classified as (a) analytical (direct observation, whether during an interview or in a partially structured group situation; questionnaires, almost exclusively American; and objective tests of personality), and (b) syncretique or projective (their rationale and an inventory of methods, following L. K. Frank's system of classification). Frequently, a technique is simply mentioned and the reader is referred to the original publication for description. The French reader will seek more detailed information in the author's recent book on mental tests in psychiatry (*Les tests mentaux en psychiatrie*: I. *Instruments et méthodes*. Paris: Presses Universitaires de France, 1949).

Stoetzel devotes 40-odd pages to methods for the study of opinion, with references limited almost entirely to the American literature. He notes that in French psychological literature the appearance of the very word "opinion" is very recent, missing in the large *Traité de Psychologie* edited by Georges Dumas and used in Lalande's comprehensive *Vocabulaire de la Philosophie* (1932) only in a non-technical meaning. Of interest is the analysis of the reasons for a delay in the cultivation of this field, especially in France: Cartesian tradition for which only "clear and distinct" ideas merit the scientist's consideration; biological orientation of scientific psychology; lack of appreciation for the role of social interaction; disregard of value judgments; domination by the theoretical models of physics; and concern for function rather than content. Stoetzel also brings up the point that institutional changes are likely to result in anxiety to which the threatened groups will respond aggressively, treating the *study* of the changes in the same way as they react to the prospect or presence of the changes themselves. The three chapters of the section deal with methods (attitudes of an individual, opinions of a group), some results obtained by the application of these methods, and the historical development of opinion research. For a more detailed treatment one should consult the author's other publications (*Théorie des Opinions*, 1943; *L'Etude Expérimentale des Opinions*, 1943; *Les Sondages d'Opinion Publique*, 1948).

Statistical methods used in applied psychology were described by Faverage. The chapter is brief (pp. 259-302), especially in reference to its ambitious content covering, as it does, both analysis of variance and factor analysis. The presentation is limited to concepts and formulae. Computational procedures are not given in a workable detail.

The potential user of the treatise would do well to arm himself with the *Vocabulaire de la Psychologie* (dictionary of psychology), published by the Presses Universitaires de France in 1951. Otherwise, he will look in vain in standard dictionaries for an English equivalent or the meaning of numerous terms (such as "docimologie," the scientific study of scholastic and other examinations). The heavy impact of the English language on the terminology of French scientific psychology—with some expressions taken over literally (tapping, closure, the term *test* itself), others in translation (*échantillonage aréolaire* for "area sampling," *sondages d'opinion* for "public opinion polls," *ensemble* for "population")—is likely to make an occasional use of the dictionary mandatory for the French student as well.

JOSEF BROŽEK.

University of Minnesota.

GILBERT, JEANNE G. *Understanding old age.* New York: Ronald Press, 1952. $5.00.

In today's society, physicians, psychologists, nurses, and social workers are called upon to solve the practical problems presented by older people and their relatives. Each of these professions has acquired a vocabulary and a body of information peculiar to itself. Furthermore, biologists, biochemists, physiologists, psychologists, and sociologists who have studied problems of aging have described their results in the technical terminology and within the concepts of their own scientific discipline. It is high time that workers from each profession concerned with aged people be offered a summary of our present knowledge of aging as gleaned from all scientific disciplines written in intelligible language. The preparation of such a volume is a formidable assignment to which Dr. Gilbert has addressed herself.

The first part of the book is devoted to a cataloging of physical, physiological, endocrine, emotional, intellectual, and social-relation changes with increasing age. Findings are presented in serial fashion with little selection for significance in terms of behavior. The reader is offered little help in assessing the importance or meanings of the changes reported. For instance, the occurrence of senile plaques in the cerebral cortex of individuals suffering from senile dementia prior to death is discussed (p. 24), but the work of Malamud showing that there is no relationship between the incidence of senile plaques in the brain and impairment of mental function is not mentioned. The reader is not informed (until p. 293) that in many cases of senile dementia, the structural changes in the brain are not

sufficient to account for the psychosis. Considerable emphasis is placed on the importance of diet in age changes, but no mention is made of the important work of McCay on the relationship between dietary restriction and longevity. In the discussion on heredity, no reference is made to the important work of Raymond Pearl. Although the Third (1952) Edition of *Cowdry's Problems of Ageing* was probably not available at the time the manuscript was written, it is surprising to find that all references to Cowdry are to the First (1939), rather than to the more complete Second Edition (1942). Neither will the professional worker be particularly impressed by the inclusion of Ripley (Believe It or Not) as a bibliographical reference. The treatment of technical terms is not consistent, since many of the simpler ones are defined (thrombosis, lumen, epithelium), whereas others are not (colloid of the thyroid gland; "hold" and "don't hold" tests). The section on intellectual changes seems preoccupied with the concept of deterioration and fails to emphasize the findings that decline is minimal or even absent in intellectual functions that are practiced and used in the personal life of the aging individual. In discussing age changes in creative ability, the author leans heavily on anecdotal material rather than on the careful studies of Lehman and his co-workers.

Part two of the book deals with abnormal life changes in aging, but no clear definition of abnormal is made. Actually this section deals with clinically diagnosable disease states found in elderly people. Physicians will no doubt look askance at the definite statements of symptoms and methods of treatment of diseased states. A good many errors of fact

have crept into the presentation. Although there is a great deal of theorizing about emotional factors and their causal relations to disease states, the author leads the unwary astray in presenting these theories in the same way as factual observations.

The third section of the book, concerned with professional work with the aging, contains a great deal of common sense presented in a readable fashion. The section is specific and offers many valuable suggestions on the theme that old people can be integrated into effective community life. The blind acceptance of the value of vitamin supplements is a limitation. Of special value is emphasis on the idea that although we as yet have no panaceas for reversing the physical changes of aging, much can be done on the emotional and personal reactions of individuals toward these changes.

Nathan W. Shock.
Baltimore City Hospitals.

Reik, Theodor. *The secret self.* New York: Farrar, Straus and Young, 1952. $3.50.

Theodor Reik here continues the process, elaborated in *Listening with the Third Ear,* of being both the psychoanalyst *and* his couch. In the present volume he gathers some of his early ruminations on diverse topics and expands upon them in free-associational contemplation. The points of departure range from lamb chops on a restaurant menu and thoughts on Goethe during an analytic session, to hypnagogic interpretations of parent-child relationships within Reik's family.

The psychoanalyst may look with favor on the richness of commentary developed around each leitmotiv, but the scientifically oriented reader will look askance at the methodology and its rationale. Juggling phrases, al-

lusions, and emotional content, Reik attempts to explore the unconscious motives of the heroes, authors, or prime-movers in each of his vignettes. Unfortunately, the phrases, allusions, and emotional content are not those of the heroes and authors, *but of Reik himself* as he responds associatively to clues which perhaps he alone perceives. What justifies this approach? To quote Reik (e.g., page 170), "These impressions are elusive and allusive, too, but the analytic technique will, I hope, help us in attributing a new significance to them, and in evaluating them in the service of psychological understanding."

The reviewer was tempted by this justification to interpret in similar fashion the handy and obviously Freudian slip occurring in the chapter on *Sexual Symbolization in Modern Plays.* On page 225, Reik declares (or was it the typesetter?): "It is possible to speak of 'emotional uticaria' or puritis." Freely associating, the reviewer perceives an error common to both medical terms used—the omission of the "r" in each. The connotation is immediately clear —it is reminiscent of those days when he (and obviously Reik, too) was frustrated by the lack of edible oysters in the devastating "r"-less months of the year. The culturally determined sexual symbolization is seen to be thematically related to Brocchi's *Traité de l'Ostréiculture. . . .*

The book evokes—this gentle scoffing notwithstanding—a genuine admiration of the cultural erudition which Reik brings to his analyses, an erudition applied with practiced familiarity and without ostentation. It is with regret that one must, in the face of this, accuse Reik of a major modern crime—guilt by free association.

Dan L. Adler.
San Francisco State College.

Psychological Bulletin

IMPROVEMENT IN PERCEPTUAL JUDGMENTS
AS A FUNCTION OF CONTROLLED PRACTICE
OR TRAINING[1]

ELEANOR J. GIBSON[2]
Cornell University

In spite of a long history of research on problems of learning, current textbooks contain little information relating perceptual judgments to learning. That perception is educable is clear from many "real life" instances. A butcher can heft a piece of meat and estimate its weight with great accuracy. The physicist Michelson (**198**) became so skilled at estimating the degree of distinctness of the fringes of light waves that these visibility ratings permitted him to compute the nature and shape of the spectrum lines. The "facial vision" of the blind has been shown to be attributable to their learning to perceive and respond to minute changes in the reflected sounds originating in their own actions (**26, 159, 203**). And there is William James's eloquent statement of the case. He said: "That 'practice makes perfect' is

notorious in the field of motor accomplishments. But motor accomplishments depend in part on sensory discrimination. Billiard-playing, rifle-shooting, tight-rope dancing demand the most delicate appreciation of minute disparities of sensation, as well as the power to make accurately graduated muscular response thereto. In the purely sensorial field we have the well-known virtuosity displayed by the professional buyers and testers of various kinds of goods. One man will distinguish by taste between the upper and lower half of a bottle of old Madeira. Another will recognize, by feeling the flour in a barrel, whether the wheat was grown in Iowa or Tennessee. The blind deaf-mute, Laura Bridgman, had so improved her touch as to recognize, after a year's interval, the hand of a person who once had shaken hers; and her sister in misfortune, Julia Brace, is said to have been employed in the Hartford Asylum to sort the linen of its multitudinous inmates, after it came from the wash, by her wonderfully educated sense of smell" (**87**, p. 509).

During World War II a need developed in many areas for knowing the potential perceptual skill of a man in judging such aspects of the world as the size, distance, angle, and speed of obstacles, targets, and other

objects in the environment. Often the available level of skill left much to be desired and the question at once arose whether training could improve it. The kind of practice which will be most effective in increasing the skill of perceptual judgments is clearly a problem for experimental investigation. It is true that there has been a resurgence of interest in learning as an explanatory concept in perception, as witness Ames's revival of the Helmholtzian unconscious inference (20, 86, 96), Hebb's empirical and neurological theory (69), and the "new look" in perception, which stresses cultural relativism and individuality of perception, with past experience as a possible determiner of differences (13). Yet these new and very popular trends have not, to any impressive degree, resulted in a program of experimental research on learning in perception, that is, with experimental control of practice and experimental measurement of perception.

It is the purpose of the present review to bring together available facts which show the effects of training on perceptual judgments, with the hope that some specific questions can be answered. Can perception be rendered more acute by training, in the sense that the observer (O) comes to discriminate smaller differences in stimulation? Can perceptual judgments of single stimuli be brought into more precise correspondence with physical scales? Can the ability to recognize complex patterns of stimuli under conditions of dim light or short stimulus duration be improved? What are the most favorable conditions for perceptual learning? And how general is it? Can transfer be expected? These are all practical questions, both in the sense of having immediate applicability, and in the sense that the psycholo-

gist has the methods for answering them. The answers can be looked for in experiments which control and vary the frequency of practice and other factors such as kind and amount of reinforcement. The perceptual judgments of O, the dependent variable, can be obtained and measured by the techniques of psychophysics.

The studies chosen for report are limited by these questions to ones which deliberately manipulate practice in the experimental situation, or at least quantify practice which took place outside it.[3] This formulation of the problem of perceptual learning will be referred to as the problem of *improvement* in perception. The criterion of improvement will be defined in terms of veridical judgment, that is, evaluation of the judgment in terms of standards known to the experimenter (E) by virtue of the physical yardsticks available to him. If an O is being trained to estimate

---

[3] Other limitations should be noted. The effect of practice on judgments primarily of interest for reasons of physiological explanation has not been included; for instance, flicker fusion, though it may, apparently, show threshold changes as a function of practice (207). Experiments on animals are omitted, except in a few cases where they happen to suggest general principles, because it is too hard to disentangle the learning which is really due to perceptual practice and that which is due to learning the nature of the task and the response categories demanded by E. This latter factor, in human experiments, may play a role, but it is usually minimized by the preliminary instructions given O. Finally, it will be noted that very few experiments are reported in which O designated his judgment by a motor response other than a verbal one. This is not a systematically defensible omission, in itself, but unfortunately, the more complex the motor response required, the less clear it becomes whether any perceptual learning has taken place. All the "deprivation" experiments have been omitted—that is, experiments in which normal practice is prevented by temporary immobilization or sensory deprivation, but practice itself is not varied.

angles in a gunnery course, the measure of his learning is given by the deviation of his estimates from the true angle as defined by some measuring instrument such as a protractor. Improvement is arbitrarily defined, then, as closer, more precise, more immediate approximation of O's judgment to the appropriate physical standard or measure. This definition means that some kinds of judgment, such as Rorschach responses, are excluded, since there is no way of scaling "rightness" or "wrongness" and plotting a learning curve. Highly personalized perception, while certainly affected by O's past experience (13), is irrelevant for the problem as stated, owing to the absence of a criterion which permits measurement of learning.

Practice, for present purposes, will be defined as any controlled activity of O which involves repeated perception of the test stimuli or ones similar to them. This definition assumes attention on the part of O, but it deliberately omits any requirement of reinforcement, correction, or reward. The role of these factors will be examined in the light of the evidence. The facts which demonstrate improvement in perception will be summarized first, and then the factors which affect this learning, such as amount of practice and reinforcement, will be considered. Finally, transfer and retention of perceptual learning will be discussed, in so far as the facts permit.

EVIDENCE OF IMPROVEMENT OF PERCEPTUAL JUDGMENTS BY PRACTICE

The psychological literature of the past 70 years contains a vast amount of evidence that perceptual judgments can be improved. It will be summarized briefly, since the extensiveness of the literature does not permit inclusion of experimental details. Classification will be mainly by the operations used in measurement.

Acuity

The psychologist is tempted to think of acuity as a baseline which defines the structural limits of perceiving. Yet, acuity measures do vary with such factors as the kind of judgment required, and the lighting and nature of the test object. Possibly they vary also with practice in the judgment demanded. The effects of training have been investigated in the following areas.

Correction of visual anomalies. There have been numerous claims by optometrists that not only an improvement of acuity, but also the correction of astigmatism, eye-muscle imbalance, weak binocular fusion, nearsightedness, and other anomalies can be accomplished by prescribed training, usually involving stereoscopic targets (17, 23, 102, 131). Woods (200), at the Wilmer Institute, reported results on 103 myopic patients tested before and after such training and concluded that expected normal variation accounted for most of the apparent improvement. Morgan (127) gave similar training to patients with seven types of anomaly, and concluded that the treatment must be varied depending on the source of the difficulty. Peppard (133) recommends the "Bates" method of exercise (practice in "central fixation," relaxation, the "long swing," etc.) for correction of visual defects. The validity of his claims has been discussed by Allbaugh and Miller (1) and by Lancaster (101). They are not supported by experiment.

Foveal visual acuity. On the other hand, positive evidence has been obtained, in an experimental situation, for the effect of practice on a typical

acuity judgment. Volkmann (182), in 1863, first reported such a change. Later Sanford (148), using letters as test objects, and Wilcox (194) and McFadden (119, 120, 121), using parallel bars, found large increases in acuity with practice. Bruce and Low (12), testing with Landolt rings, reported an increase in acuity after training with tachistoscopic presentation of aircraft photos.

Peripheral visual acuity. Dobrowolsky and Gaine (31) in 1876 reported an increase in peripheral visual acuity with daily practice sessions up to six weeks. Low (114, 115) found that his Os could respond to stimulation of an area only 1/11 the size of that necessary before training, after 25 hours of practice with airplane silhouettes presented on various meridians of a perimeter. Test objects were Landolt rings of progressively decreasing sizes.

The two-point limen on the skin. Since the receptor surface of the skin is spread out like the retina, the techniques of investigating cutaneous sensitivity are analogous to methods of determining visual acuity. Titchener (167) and Luciani (116) both commented on the effectiveness of practice in lowering the two-point limen. Volkmann (181) found in 1858 that the distance at which two points were felt as double could be halved with practice. Dresslar (33), Solomons (156), Tawney (162), and Mukherjee (128) all obtained similar decisive results. Hoisington (77) contradicted this finding, but several factors, including the O's previous level of practice, which was high, and the area chosen for stimulation, the forehead, where the two-point limen is already quite low (see 81), could explain the lack of improvement.

Upper and Lower Thresholds

Like acuity, the upper and lower limits of sensitivity appear to be susceptible to the influence of practice. Titchener, comparing determination of RL's by two psychophysical methods, said, "The RL is a variable value: variable not only for different O's, at different stages of practice, with different degrees of attention, but intrinsically variable" (167, p. 21). He noted the lowering of the intensive RL for pressure with practice, and persistence of the effect after a three-day interval. That practice must be allowed for in getting such a limen is clearly indicated in the methodology of psychophysics. Guilford, for instance, in discussing the method of measuring the lower limen for pitch of tones, gave sample data for its determination by the method of minimal changes in which the difference between the first and last ten series was "almost significant enough to suggest a lowering of the limen as if by a practice effect during the course of the experiment" (65, p. 121). Humes (84) found that practice raised the upper limen for tonal discrimination. Brown (11), in studying the absolute threshold for salt with human Os, thought the limen was not progressively lowered with practice, but that it fluctuated with the range of stimuli presented and with O's attitude. But Harriman and MacLeod (68) demonstrated a very significant lowering of the salt threshold in rats as a result of practice under conditions of thirst motivation and electric shock reinforcement.

It has been shown by Verplanck, Collier, and Cotton (176) that even when O's threshold is relatively stable on succeeding days, the responses themselves are dependent on previous responses; that is, runs of "yeses" and "noes" occur which are of greater length than expected on a hypothesis of independence. The same group of investigators found a

a shift downward of as much as .10 log units in the visual threshold after occasional reinforcement by *E* during threshold measurements (177).[4] Upper and lower thresholds and acuity are probably analogous to a physiological. limit concept in learning, and do not necessarily imply a basic sensation in fixed correlation with a receptor structure. The very concept of a threshold is a statistical one; a lowered threshold as a result of practice is not paradoxical or inconsistent with any doctrine except the doctrine of fixed and immutable sensations automatically touched off by stimulation of a specific receptor.

Discrimination of Hue

Can hues not originally within *O*'s limits of discriminability become effective stimuli with training? During the war, a number of optometrists attempted to "train" hue discrimination in color-deficient *O*s and in some cases claimed success (35, 36, 109). It seems likely that improvement in these cases was due to increasing information about the tests themselves, which were usually some form of pseudo-isochromatic charts. Gallagher, Ludvigh, Martin, and Gallagher (54) found that training with such plates did not improve performance on a desaturation test, nor did training with the latter improve performance on pseudo-isochromatic tests. Chapanis (22) tested three "improved" cases after four years and found that pseudo-isochromatic charts could be read only where learned clues enabled *O* to call the correct responses. Learning correct responses for pseudo-isochromatic plates could be cued by brightness differences in the plates (9, 45).

[4] For another instance of lowering of an absolute threshold by training, see Blackwell, H. R. Psychological thresholds. Ann Arbor: Univer. Mich. Engng Res. Inst. Bull. No. 36, 1953.

Discrimination of Relative Differences

Can the discrimination of small differences between stimuli be improved by practice in comparing one with another? A relative discrimination will be taken as a judgment which consists of a comparison of the given stimulus with a designated preceding or adjacent stimulus (183), in contrast with an "absolute" estimation. In asking whether relative judgments can be improved, the distinction must be made between variable and constant errors—we want to know whether a finer just noticeable difference can be achieved with training. An *O* could correct a constant error in bisecting a Galton bar without necessarily lowering his variable error. On the other hand, DL's and constant errors could be lowered at the same time.

Pitch. The effect of training on relative discrimination of pitch has been investigated in 12 studies, only two of which found no improvement. The method of practice is undoubtedly a crucial factor here. Stanton and Koerth (157, 158) used no special practice, but gave the Seashore test before and after three years of general music study at the Eastman School. No general improvement in relative pitch discrimination resulted. Buffum (reported in 206) found no improvement with mere repeated administration of the test situation. All the others obtained lowered thresholds, although Wright (204, 205) found improvement only in the six lowest initial scorers after five repetitions of the Seashore pitch discrimination test. Smith (reported in 206), Whipple (192), Smith (154), Cameron (18), Wolner and Pyle (197), Seashore (152), Capurso (21), Connette (24), and Wyatt (206) obtained improvement in varying degrees, in some cases very impressive. Elaborate training methods caused

DL's to be halved or better. In one study, children who initially could not distinguish octaves could distinguish semitones or finer after training (197).

Weight. Brown (10), using the constant method with lifted weights, gave Os a large number of trials with correction, and found that judgments of larger differences showed irregular improvement ("lopping off of the coarser errors"), but no better discrimination of minimal differences. Urban (171) studied the effect of progressive practice in lifting weights, using a constant method, but no correction. The coefficient of precision (*h*) increased, and the interval of uncertainty grew smaller. Fernberger (42) repeated the experiment with similar results.

Visual space. The judgment of nearer-farther as a function of practice has apparently been studied only under conditions which eliminate many of the normal cues. Van Tuyl (173) eliminated all cues except accommodation and convergence and asked Os to judge which of two lights presented successively (in a dark room) was nearer. Only two of seven Os showed learning curves. Since blur would accompany change in either direction, it is hard to see how learning could occur, except by way of accompanying convergence cues. One O, who was myopic, "learned" to call the far points near and the near points far beyond a certain distance. Theories of space perception which assume that we see distance because we learn meanings for cues, such as changing accommodation, receive little support from this study. On the other hand, an experiment similar to Van Tuyl's by Peter (reported in 201, p. 672) found that one O, helpless at first, learned to utilize cues of time and difficulty of focusing. The cues

were inferential at first, but the depth impression, with practice, became immediate.

That relative judgment of two-dimensional length, or visual extent, can be improved with practice has been shown by Wolfe (196), Moers (126), Woodrow (199), Hamilton (67), Ueno (170), and Eagleson (37). Both CE and MV may be decreased in this type of judgment; whether either one or both types of improvement occur probably depends on whether correction is given. Welch (189) found that relative discrimination of form (varying along a continuum of squareness to tall rectangularity) was improved by practice in young children.

Spatial illusions provide an excellent situation for determining the effects of practice on otherwise compelling constant errors. Many experiments have investigated this question, with conflicting results. The conflict is probably due to lack of control of O's attitude; O can always compensate deliberately for an illusion which he knows is present. Judgment of the Müller-Lyer illusion became more accurate with practice in experiments by Judd (92, 93), Lewis (110), Seashore, Carter, Farnum, and Sies (150), Crosland, Taylor, and Newsom (28), and Köhler and Fishback (98, 99). Several Es reported that the illusion diminished gradually and without conscious correction. Judd reported that "it comes to look differently than it did at first" (92, p. 29). Three of the Os of Köhler and Fishback developed a "negative illusion."

Other illusions have yielded equivocal results with practice. Williams (195) found no change in the horizontal-vertical illusion, but Os "tried" not to let their knowledge of the illusion affect their judgments. Ritter

(143) studied judgments of length or lines inclined at various angles from the horizontal. The illusion tended to decrease if the line originally overestimated was varied in length during practice. Seashore *et al.* (150) obtained ambiguous results with cylinders in vertical and horizontal positions, also with a *T* illusion, and in estimating the distance between two circles. Cameron and Steele (19) found gradual disappearance of the Poggendorf illusion with long practice. Judd and Courten (95) found a decline in the Zöllner illusion with practice, after an initial rise. In these experiments, most investigators were convinced that the decrease in CE occurred without any conscious attempt at correction, but O's attitude was not deliberately varied in both directions in any study.

Absolute Estimation and Rating

That absolute judgments of stimulus magnitude and quality may be improved with practice has long been recognized. The method of "single stimuli" is admirably adapted to the study of "object judgments" or ratings, and the literature in this field has for a long time taken account of O's experience with the test stimuli. Hollingworth's concept of "central tendency of judgment" (78) and Helson's "adaptation level" (71) reflect such an interest, as does the literature on anchoring (122, 145, 184). There is a correspondence between the actual range of stimuli presented and O's conception of the stimuli as a series; this conceptual series tends to persist, but is responsive to change in the range of stimuli presented (130, 191). The category thresholds and the midpoint of the scale will reflect such a change. Johnson (88, 89) has incorporated these facts in a theory of how a sub-

jective scale is learned. The units (categories) of the scale may be familiar and conventional ones, such as inches or minutes, or they may be arbitrarily chosen by *E*; the number of response categories may vary, and they may be adjectives, such as "heavy" or "light," or numerical values. He makes the following assumptions: that the scale is related to physical values of the stimulus objects presented; that each judgment of a stimulus object constitutes a unit of practice; that *O* makes his judgment with reference to the midpoint of the scale; that any given stimulus generalizes up and down the scale in linear fashion; that the midpoint of the scale is the arithmetic (in some cases, geometric) mean of the central effects of the stimuli upon which the scale was founded. These assumptions are nearly all accepted facts of judgment; the novelty lies in Johnson's attempt to relate scale development to learning theory by counting units of practice. He combines these hypotheses in equations yielding predictions which he has checked with weights (88) and with tones differing in pitch (90). His results show, in general, good agreement between predicted and obtained limens. A further experiment with pitches (91) investigated the change in the scale with introduction of new stimuli offered for judgment. The midpoint of the scale was found to shift, as predicted, and furthermore, the shift was retarded relative to the frequency of preshift trials with the first series. Tresselt (168) anticipated the latter finding in an experiment with weights in which she found that as practice with the first scale increased, the scale of judgment tended to shift more slowly to its new position. The "range" effect can be related to *improvement*,

since the subjective midpoint turns out to be anchored to the physical midpoint of the series used. If O has experienced the entire range of relevant stimulation (all that the environment provides), his midpoint would be realistically anchored.

The training procedure most often followed in the studies providing evidence of improvement in absolute estimation is quite similar to a paired-associates procedure. The stimuli to be judged are paired with labels indicating the appropriate magnitude on the physical scale chosen. The difference between this learning situation and ordinary identification learning (learning names of people, things, nonsense forms, etc.) is that the range of stimuli may be placed along a dimension or continuum; furthermore, the responses which O is to learn may be scaled and placed in a consistent relationship or correlation with the stimulus continuum. The term "scale training" might be used to indicate the procedure. The following experiments have demonstrated that improved judgments result from such training.

Pitch. According to Petran, "Judgments of absolute pitch may be defined as judgments based on associations learned between more or less narrowly limited ranges of the pitch series and the terms of any unequivocal nomenclature, these judgments being without reference to or aid from any tone or tones recently heard which have been given as a standard or attended to in any degree as being of a certain pitch or familiar pitch position" (134, p. 12). The O starts with a large number of very similar stimuli and gradually achieves identification of each tone by its appropriate name. But precision, even for a highly trained musician, is never complete. Petran

showed that there is a narrow range of equivalent frequencies, rather than a single numerical frequency which is adequate to elicit the judgment of a given note.

Four studies have dealt specifically with the effects of practice. Meyer (124) used progressive practice with a total of 39 notes, a whole tone apart. Notes were called by vibration rates read from a chart. Four months of practice produced improvement, which was largely lost after several years of no practice. Accuracy was inversely related to the number of tones practiced. Gough (63) used all 88 keyboard notes on the piano; O was instructed to name each note and its octave from a keyboard model. Varying procedures were used. Average error for all Os decreased from 5.5 to 4.5 semitones. Those who practiced longer improved more. Retention was good after a year. Practice on a single note improved recognition of it. Fewest errors occurred in the middle octave, perhaps because of more frequent practice of this octave. Mull (129) trained Os on only one note and tested them by requiring recognition of this note among a group of nine tones. Average error decreased from 2.85 to .33 semitones and finally to .29 semitones when the difference between the tones was narrowed. Wedell (187) had Os identify tones by vibration rate, represented at appropriate points on a chart. Average error in DL's was cut in half or more by training. Greatest error occurred in the middle of the scale. There was no tendency to constant error. With progressive practice, all Os learned to identify nine tones without error. When the series was longer (13, 17, and 25 tones), no one mastered it in the number of trials given, but a learning trend is apparent.

Weight. Fernberger (43) had *O*s judge weights on an absolute three-category scale. Practice had the same effect as in a relative series; precision increased, the interval of uncertainty decreased, and the *p.s.e.* increased. Practice also resulted in a reduction in reaction time (Fernberger and Irwin, 44). Thorndike and Woodworth (164) gave *O*s training in estimating weights in grams and found improvement.

Visual space: area and extent. The classic experiments of Thorndike and Woodworth (164) show that the estimation of the areal size of different geometrical shapes, such as rectangles, circles, and triangles, can be improved. In another experiment, *O* judged lines of different length, and improvement was again evident. One gets "more accurate mental standards and more delicacy in judging different magnitudes by them. In the case of estimations of magnitudes in terms of unfamiliar standards such as grams or centimeters, the acquisition of the mere idea of what a gram or centimeter is, makes a tremendous difference in all judgments" (164, p. 394). Thorndike (163) later did further experiments on visual estimation of length in centimeters and area in square inches, with marked improvement resulting.

Van Voorhis (174) conducted a large-scale experiment on improvement of "space perception ability," the criterion being score on the Cards and Figures section of the Thurstone test for Primary Mental Abilities. Training included estimation of linear extent, angles, and areas, various visualizing exercises, three-dimensional tick-tack-toe, and practice with a stereoscope. Scores improved significantly relative to a control group.

Postman and Page (138) had *O*s judge either height or width of rectangles by a variant of the method of single stimuli. Precision increased during the series and DL's decreased. Retroactive inhibition resulted from interpolation of the alternative judgment.

Visual space: angle. In an experiment done in an Air Force Gunnery Training program (76), gunners were trained in estimating the "aspect angle" of a plane. This was the angle between the line from gunner to target and the line extending through the longitudinal axis of the target aircraft. A training device was used with a model fighter set at a simulated range of 400 yards and variously inclined to the line of sight. Judgment was improved by training.

Visual space: depth and distance. Judgments of stereoscopic depth improved with practice in two experiments where stereoscopic range finders were used (79, 209). Range-estimation trainers which require a judgment of horizontal extent or of proportion of the reticle filled have also been studied. An experiment of this type with the RCAF range-estimation trainer found marked improvement as a result of even a small number of training trials (76, pp. 148 ff.).

Studies of the effect of training on unaided visual-range estimation have yielded positive results (80, 180, 208, 210, 211). Two large-scale studies used aerial targets viewed against an expanse of sky (80, 179). The *O*s estimated in yards; range of targets varied, in one case up to 8,000 yards. The *O*s improved with corrected practice in both cases, and constant errors were generally reduced. A comparison of various methods of training for estimating a fixed opening range (144, 178, 185) suggested that training on the firing line was more effec-

tive than special trainers, and that unaided vision, after training, was more accurate than stadiametric estimation. The Princeton Fire Control project (210) utilized existing targets, such as telephone poles viewed over a ground surface, at distances varying from 235 to 5,940 yards. Training resulted in improved accuracy of estimates, and also in a reduction of variable error for both the group and individual Os. Constant errors shifted irregularly for different targets.

Visual speed. Speed of motion of airplanes of various types, flying different courses at different speeds, was estimated by anti-aircraft officers in an experiment by Biel and Brown (6). Corrected practice had the effect of raising estimates all through the speed range, causing slow speeds to be overestimated more than before training, so that improvement was limited to high speeds. Group *SD*'s decreased, but consistency of individual estimates was not improved by training.

Visual form. A study was made by Gilliland (61) of the effects of practice on grading handwriting. Unaided practice in grading samples on a 1–100 scale was followed by practice with a "true" scale for reference and with knowledge of results. Errors were greatly reduced.

Brightness and hue. Though no direct study of training has been made for absolute estimation of these qualities, two studies are suggestive. Lehmann (108) showed Os a series of grays arranged in a scale from black to white. Then each member was exposed singly and O was required to judge its order in the series. A five-member series was judged with great accuracy (96.7%), but with longer series accuracy dropped. Lehmann thought that the absolute scale was

limited by O's ability to label the stimulus presented, and that O had labels readily available for only a five-member scale (black, dark gray, medium gray, light gray, white). Halsey and Chapanis (66) found that the number of absolutely identifiable spectral hues was limited to 10 or 12, even after practice, but they expected that certain errors might be eliminated with further practice.

Kinesthetic extent. Thorndike (163) showed that Os improved in their ability to draw a line of a specified length, such as three inches, while blindfolded, under certain conditions of practice. The result has been corroborated (169).

Pressure. Thorndike (163) also reported experiments on gauging pressure of a specified strength on a dynamometer. Practice followed by "right" or "wrong" resulted in improvement.

Recognition of Patterned Stimuli under Impoverished Conditions of Stimulation

When patterned stimulus material is presented under conditions deliberately made far from optimal, such as very low illumination, brief exposure, presentation to parts of the sensory surface where receptors are few, reduced size, etc., the conditions are referred to as "impoverished" (58, p. 166). Test objects which might be easily identified under favorable conditions may not be recognized. A threshold measure can be calculated in terms of the amount of light, duration of exposure, etc., required for accurate identification. Many experiments have investigated the effects of practice in perceiving under these conditions.

Peripheral presentation. A series of experiments on "peripheral retinal

learning" were conducted by Franz and various collaborators (**48, 49, 50, 51**), all of which showed improvement in accuracy of recognition of geometrical forms or words with repetition of them. Three figures were exposed at once, for .1 sec., displaced respectively to the right, left, and above a fixation dot. Drury (**34**) also studied the effect of repetition on peripheral perception, and found that drawings of the diagrams presented became stabilized, though not necessarily more veridical. Henle (**73**) tested accuracy of peripheral perception as a function of frequency of experience with the figure *prior* to the test situation, and found that past frequency facilitated perception. Postman, Bruner, and Walk (**137**) confirmed the result.

Low illumination. Seward (**153**) studied the effects of practice on identification of letters presented dimly on a ground-glass screen for 1.5 sec. each. Practice (without correction) resulted in gradual, continuous improvement of all *O*s. Bevan and Zener (**5**) found thresholds for recognition of nonsense forms by raising illumination to a liminal level and showed that both general practice effect and specific familiarity affected the threshold, the specific practice effect being greater and proportional to frequency of pretest exposure.

Recognition of ships or obstacles by night lookouts is a practical problem similar to the studies just mentioned. Training programs for night lookouts were established by the armed forces during World War II. Training usually took the form of practice in looking at models and diagrams displayed in dim light in a viewing box (**172, 188**).

Brief presentation. Repeated presentation of the same stimulus material in a tachistoscope has been the traditional approach to what is often called the "development" of perception. Older studies were primarily concerned with the stages of development, but it can also be shown that repeated brief presentation may result eventually in accurate perception. Fehrer (**41**) presented line figures successively in short exposures. Stages occurred, but the figures were eventually perceived correctly. Irwin has shown that limens for accurate recognition of barely perceptible patterns presented for .18 sec. are reduced with practice (**85**). Postman and Bruner (**136**) investigated the effect of different kinds of preliminary training on the threshold for perception of a gap in the outline of a circle. Preliminary exposure of open circles reduced the threshold for perception of an opening. But previous viewing of closed circles also reduced the threshold relative to a control group, though not as much.

It is possible to study the effect of practice on the *amount* of material which can be reported after tachistoscopic exposure. Historically, this is the problem of the "span of apprehension" as a function of practice. Tinker (**165**) reviewed the pertinent experiments in 1929 and concluded that practice increased the range of apprehension, but that the effects were very specific. A later experiment by Weber (**186**) corroborated his conclusion. Educators took up the same problem under the heading of "flash training," and their work will be considered under transfer, since transfer to reading skill was the question at stake. But the results uniformly showed that training with a tachistoscope has a significant effect on performance in the training situation (**55, 57, 142**).

The judgment of number as a func-

tion of practice has been studied in three experiments. Taubman (161) found that practice in estimating the number of tones presented at rates of .1, .07, or .056 sec. resulted in improvement, though overcompensation for a constant error sometimes occurred. Saltzman and Garner (147) had Os estimate the number of circles in a group of concentric circles exposed for .5 sec. With practice, reports became more accurate and reaction time decreased. Knowledge of the range of the number of circles improved accuracy immediately. Minturn and Reese (125) studied the effect of differential reinforcement on estimation of number of dots presented visually for .2 sec. Error decreased, though there was some tendency to overcompensate for constant errors.

Relative familiarity with material gained prior to the tachistoscopic viewing also affects differentially the accuracy of perception. Vernon (175, pp. 123 ff.) showed that unfamiliar words were misread when presented tachistoscopically 52 times each, compared with 2.8 times for more familiar words. Leeper (107) familiarized Os with Street figures (incomplete pictures) by preliminary exposure, naming, demonstration by E, etc. In a later tachistoscopic test (.01 sec. exposure), these Os named the figures more accurately than a control group. Howes and Solomon (83) found a strong inverse relationship between relative word frequency (frequency of occurrence in the English language) and visual duration threshold. In the same experiment, the effects of tachistoscopic training are also apparent. The frequency-duration threshold relationship has been corroborated by Postman and Schneider (139) and by McGinnies, Comer, and Lacey (123). Recency of word usage has also been correlated with duration threshold (140). Differential word frequencies which were established experimentally have also been shown to be correlated with duration threshold (155).

Noise and distortion. The problem of communication against a background of noise is in some ways comparable to viewing under impoverished conditions. When listeners were trained for hearing over an interphone system (7), greatest gains were produced by practice under test conditions with the actual words used. But practice with the words under other conditions helped. Howes (82) analyzed data collected by Mason and Garrison (118) on intelligibility of spoken messages heard under conditions of noise and found a significant correlation between frequency of correct transcription and average frequency (probability of occurrence) of the words used. That improvement in understanding of masked or distorted speech occurs with practice over and above increasing familiarity with the words spoken has been shown by Licklider and Pollack (111) and by Egan (38).

FACTORS INFLUENCING IMPROVEMENT

Amount of Practice

That improvement in perceptual judgments occurs with practice is evident. What is the function which describes the relation between amount of practice and degree of improvement? Merely to say that there is an effect does not satisfy the psychologist interested in learning; he wants to see the learning curve. Undoubtedly, there will be many curves, not one, depending on the operations measured, units chosen,

and other factors. And unfortunately, few of the experiments have measured the effects of practice at enough points along the baseline to describe a function. That frequency is a significant variable can be concluded, however, from a number of studies (5, 28, 41, 63, 83, 153, 168, and others), and that early practice is more effective than later stages is probable (Fernberger, 42).

Learning curves have, it is true, been plotted in a few studies. Woodworth (201, p. 182) plotted data obtained by Volkmann (18) for the effect of practice on the two-point threshold. The curve of errors falls gradually with a negative acceleration. Seward (153, p. 33) presents individual curves for the number of letters correctly reported under conditions of dim illumination and brief exposure as a function of days of practice. All the curves show a very gradual and continuous rise with no marked acceleration. Howes and Solomon (83, p. 407) plotted duration thresholds showing the effect of amount of practice on tachistoscopic viewing. About three-fourths of the practice effect was accomplished in one-fourth of the experiment, but the drop is continuous and is still apparent at the final (60th) threshold measured. Bevan and Zener (5, p. 439) plotted the curve for the intensity threshold required to perceive meaningless figures as a function of pretest frequency of viewing. The decrease in intensity is progressive and negatively accelerated.

In contrast to these gradual and continuous functions, Minturn and Reese (125, pp. 222 ff.) found that reduction in error of judgments of numerosity occurred very suddenly. In this case, Os were able to compensate immediately for constant errors which had been consistently

present in pretraining judgments. But correction of a constant error is not always immediate, for Judd (93) found that 24 days' practice with the Müller-Lyer illusion yielded a gradually sloping curve (average error dropped slowly from 17.3 to 1.7 mm.). The O was said to know the effects of his practice.

Reinforcement

Ample evidence exists to establish the proposition that improvement in perceptual judgments is a function of frequency of practice. Does this mean that frequency of repetition, pure and simple, is a sufficient condition for perceptual learning? Or must practice be reinforced in some way?

Perceptual learning without reinforcement. Evidence of perceptual learning without any *apparent* reinforcement does exist in experiments where E has not deliberately introduced correction or reward. Studies of the two-point limen in at least four separate experiments (33, 128, 162, 181) showed learning without reinforcement by E. However, a method of limits, as is customarily employed in these experiments, provides O with a clear example of both "twoness" and "oneness." Perhaps he can check himself, to some extent, by these anchoring experiences. The limen did not decrease when a constant method was used (77). In experiments on absolute judgment by the method of single stimuli, merely "getting the range" of a set of stimuli can affect O's judgments (89, 130, 191) and stabilize them in predictable ways. No reinforcement is given by E, though O may predict and check for himself as his concept of the range is anchored by experience with the stimuli. Woodworth took the position that there is a direct perceptual motive "to see clearly, to

hear distinctly—to make out what it is one is seeing or hearing" (201, p. 123). Such a motive would lead O to reinforce himself insofar as possible. But several experiments appear to present almost no opportunity for even self-reinforcement. Relative comparisons of lifted weights (42, 171), for instance, improved with practice, although no knowledge of results was given. Successive presentations with a tachistoscope (41), and practice in viewing letters under dim illumination (153), resulted finally in accurate identification of stimuli. Reinforcement could have played very little role in these cases.

It is possible that judgments become more consistent in the absence of external reinforcement, without becoming more veridical. In Drury's experiment (34) on repeated peripheral presentation of patterned stimuli, all Os become more consistent in their drawings of the stimuli, but not necessarily veridical. When stimulation is ambiguous, O's perception is apt to become schematized and stabilized, as Bartlett (4) showed, but veridicality probably depends on an external check. Seashore and Bavelas (151), using Thorndike's data (163), showed that repeated practice in drawing a line of a given length, without any correction by E, resulted in O's closer and more consistent approximation to his own standard. They thought this stabilization was due to O's *presumed* knowledge of results, however faulty.

If Woodworth's theory of perceptual reinforcement is correct, O's perceptual judgments of the spatial world around him should tend toward veridicality, rather than internal standards, because locomotion in this world could provide the external checks needed.

Effect of reinforcement. Although

the possibility must be admitted that perceptual learning may occur without correction or reward, all those experiments which provide a contrast of practice with and without knowledge of results show a clear superiority of the condition with knowledge. Thorndike (163) concentrated on this problem and produced evidence with many types of absolute judgment, such as visual length and area, kinesthetic length, and pressure to show the necessity of knowledge of results for improving these estimates. The E always said "right" or "wrong" in correcting O. Absolute judgments of number, practiced with and without differential reinforcement, have also been shown to improve more with reinforcement (125, 161), though Taubman (161) found some improvement without it. Even the two-point limen was lowered more effectively when E gave correction. Solomons (156) gave stimulation with one and with two points in irregular alternation. An O who was corrected lowered his limen, but another O was uncorrected and showed no change. This O did improve, however, when E later introduced correction.

Relative judgments, likewise, benefit by correction. Improvement of judgments of relative pitch was found in every case where E provided knowledge of results (21, 24, 152, 197, 206), but little or no gain occurred with mere repetition (204, 205, Smith and Buffum in 206). In comparisons of length of line, some improvement has been found to result from mere repeated practice (126, 196), but knowledge of results increased the effect. Moers (126) found that practice without knowledge decreased the MV, but actually enhanced the CE. Practice with knowledge reduced both.

The kind and amount of reinforce-

ment which will be most effective has received some attention. Probably improvement increases with the *amount of information* given by E in the correction. Trowbridge and Cason (169) repeated Thorndike's line-drawing experiment adding a condition in which O was told the amount and direction of his error. This procedure was far superior to "right" or "wrong." A similar comparison was made in an experiment by Franz and Morgan (50) on learning to identify peripherally presented forms. In three experimental conditions, O either paid attention to the forms but was not asked to report on them during practice, or he drew them as practice progressed and was told right or wrong, or he had a recognition test in which he chose the form just presented from the total group and was corrected. Learning varied as expected, with most learned in the last-mentioned condition, and least in the first. Hamilton (67), in an experiment with the Galton bar, compared conditions with (a) no knowledge of results, (b) "punishment" (a bell rang if O increased his error over a previously set standard), (c) "reward" (bell for reduction of error), (d) punishment plus a guess as to direction of error, (e) punishment plus knowledge given O of direction of error, and (f) knowledge given O of direction of error, but no bell. Improvement occurred for all conditions but the first. Improvement was least for knowledge without "reward" or "punishment." Knowledge in addition to "punishment" did not help. Hamilton thought the bell functioned as an incentive, which raises the question of the nature of the reinforcement in these studies. Is it merely informational, or is there an affective component?

Informational vs. *affective* rein-

forcement. In the studies so far reviewed, reinforcement has been generally synonymous with correction— knowledge given O by E, or knowledge gained by O from his own performance. A good example of the latter would be pitch training in which O sings the tones to be compared. In one case (206), a stroboscopic technique permitted O to see whether the tone he produced was accurate and to correct it when it drifted, thus providing a maximum amount of feedback, or information from O's performance. If the informational aspect of reinforcement is the critical factor, it might be inferred that giving the information by an anchoring technique should be as effective as correction following judgment. Anchoring was effective in bringing about improvement in a few studies (125, 138, 147), but no systematic comparison with correction has been made.

The term "informational" may be misleading, for it is possible for E to reinforce a wrong judgment. In several experiments, E has applied reinforcement in such a way as to produce perceptual illusions (16, 39, 177). The O's judgments in these cases tended to change in a predictable direction, but were in no sense "improved." Affectional reinforcement which gave no information relevant to the perceptual judgment measured was used by Lambert, Solomon, and Watson (100). Children associated poker chips with a candy reward in a two-stage token-type task. When the children reproduced the size of the token by means of a variable circle, it was made larger after the token-candy sequence, and decreased again following an extinction series. Why overestimation is produced by this technique is not clear, but at any rate a constant error is enhanced.

Whether this is learning in the same sense as the examples cited in the section on improvement is a disputable question. It seems safe to conclude that reinforcement by external correction or check is a very significant, if not an essential, variable for improvement in perceptual judgments.

Distribution of Practice

Only one study of perceptual learning, Lewis' on the Müller-Lyer illusion (110), has actually varied this parameter. Distributed practice (on alternate days) accelerated the rate of decrease of the illusion. Several studies have found, however, that periodic reinforcement is valuable for keeping the judgment at its peak of improvement. Evans (40) trained micrometer readers by having Os measure gage blocks and then read the size marked on them, noticing the size and direction of the error. When practice was continuous, there was a steady decrease in errors, but periodic practice with knowledge was necessary to maintain this improvement. Horowitz and Kappauf (80) trained Os in absolute estimation of distance. After 60 days of no practice there was an increase in average error, but a short training period restored skill. Periodic reinforcement was found necessary, also, to maintain skill with certain range-finding instruments (79).

Sequence of Practice

What is the best arrangement of material to be discriminated, if practice is to be most efficient? If the stimulus material can be scaled along a dimension, should practice be ordered in some relevant way? Common sense suggests that relative discriminations might be aided by beginning with points far apart on the scale and progressing toward closer and closer stimulus pairs. Animal experiments have in many cases borne out this hypothesis. Pavlov (132) used this method in establishing difficult discriminations in dogs. Köhler (reported in Koffka, 97), in a color discrimination study with apes, found that in a series of hues lying between red and blue, ABCDE, two neighboring ones were not originally discriminable (e.g., BC). But if the animal was given practice in discriminating a wider interval (e.g., BD), the difficult discrimination could be learned. Lashley (103) found similar transfer to a brightness discrimination; rats trained first with black and white transferred easily to a pair of grays. Lawrence (106) trained rats to discriminate shades of gray with three different methods: (a) all practice (80 trials) given on the same two middle grays to be finally discriminated; (b) 30 trials on a black-white discrimination and the rest on the two grays; (c) progressive training from black-white through a transition series to the two grays. Efficiency of learning was in reverse order for the three groups. Harriman and MacLeod (68) also successfully used progressive reduction to bring about lowering of the salt threshold in rats. That the *progressive* narrowing of the difference is important seems clear from a study by Grether and Wolfle (64), who varied pairs of brightnesses presented for discrimination by rats, but always kept the ratio of light-dark the same. Varied stimulation did not increase efficiency of learning to make a light-dark discrimination; neither did it lower it until the number of different pairs was increased to four.

But with human Os, when preliminary instructions can make clear the dimension to be discriminated, will a progressive method have any value? It is possible that the method

is facilitative in animal experiments simply because it helps the animal to discriminate the relevant dimension of difference. Progressive training may serve some other function also, however, if it makes practice more effective with human *O*s. With relative discrimination of weights, Urban (171) used progressive practice and found lowered DL's. In the relative discrimination of pitch, progressive practice was a valuable technique (152). Wedell (187), in training absolute pitch, started with five widely spaced pitches, and gradually interpolated more in the intervals between them. These cases have not provided statistical proof of the superiority of progressively ordered training; nevertheless they suggest its value. If the aim of training is to establish a conceptual scale which is realistically anchored and well differentiated (in absolute estimation of spatial dimensions, for instance), ordered practice may be the best procedure.

Role of Names and Labels

William James, in his discussion of learning fine discriminations, suggested two ways in which he thought such learning might take place. One was the progressively ordered practice just discussed; the other was learning distinctive associates to originally barely discriminable pairs of stimuli. "First, the *terms* whose differences come to be felt contract disparate associates and these help to drag them apart .... The effect of practice in increasing discrimination must then, in part, be due to the reinforcing effect, upon an original slight difference between the terms, of additional differences between the diverse associates which they severally affect" (87, pp. 510 ff.). A similar hypothesis is expressed in Dollard and Miller (32). Evidence in its favor

exists in the studies of relative pitch discrimination. Capurso (21) had *O*s learn "mood words" for intervals, and finally substituted the technical name for the mood word; Smith (154) used various associates, including adjectives (e.g. "dull," "heavy"), motor responses, and visual localization of tones in space. Three *E*s (18, 197, 206) had *O*s sing tones or intervals, thus adding kinesthetic feedback to the auditory stimulation. Learning was very effective, as exhibited by lowered DL's, in all these studies. However, they were not designed to isolate the effect of the verbal or motor response, so their role is not proven.

No truly critical experiment yet exists to test James's hypothesis with psychophysical measurements of sensitivity before and after the training. There are a few experiments now available (3, 53, 104, 105, 146) in which preliminary learning of differential responses for a range of complex stimuli has been shown to transfer to a second learning task, which involved different responses but the same stimuli, in such a way as to lower errors of generalization in the new task. But this may involve no shift in limens. That is, it is not clear whether the actual perception of differences between the stimuli has been affected. An experiment by Arnoult (2) is more relevant, but negative. He gave *O*s training in learning verbal associates (numbers) to nonsense forms. Then *O* made discriminative responses to the forms by pressing an appropriate key. Accuracy and latency measures showed no advantage over a control group which did no preliminary learning.

TRANSFER

Whether perceptual learning will transfer to new situations is a question of practical importance. Train-

ing aids, for example, must be evaluated for transfer from the trainer to the real situation, for trainers sometimes teach O to respond in terms of cues which are specific to the training device and irrelevant to the operation actually required.   For instance, Viteles and others (144, 178) studied improvement in range estimation as a result of training on the Mirror Range Estimation Trainer and found it far less effective than unaided training on the firing line, owing in part to O's tendency to respond to stimuli peculiar to the trainer. The doctrine of "formal discipline" in education is another practical case. It was first attacked experimentally in Thorndike and Woodworth's classic studies of perceptual learning and transfer (164). The O was given practice in estimating the area of rectangles; he was then given tests including rectangles and other figures. Transfer did occur, but while all six Os reduced their errors in the case of identical figures and areas, only five reduced them for the same shape but different areas, and only four in the case of different shapes. Similar results were obtained throughout: in judging the extent of lines of different length, improvement was greatest with the lines actually used in the training series, though there was some improvement with lines of other lengths. As every one knows, they stressed the importance of "identical elements" in transfer.

These results, however widely accepted, did not lay the ghost of "formal discipline," for there continued to be claims that the "power of attention" could be trained and that perceptual "skill" could be increased by practice in tachistoscopic viewing. Whipple (193) and Foster (46) investigated these claims and agreed that practice effects were specific and that transfer occurred only in proportion to similarity of the tasks. Weber (186) found no transfer from tachistoscopic training with letters to other material, and concluded there was no "G factor" of perceptual skill.

The interest of educators in "flash training" as an aid to reading skill has a similar background. Actual experiments including reading tests before and after tachistoscopic training and a suitable control group are rare; and sometimes other changes, for instance in motivation, are introduced at the same time, thus confounding the variables.   Studies which found improvement (27, 117, 149) are often of this type, and the criticism has been made by Vernon (175) and by Henry and Lauer (74). Gates (55), in a well-conducted study, found that control groups which were "reading-trained" were superior in reading to experimental groups which were "flash-trained." Other studies bearing on the problem are those of Sutherland (160), Freeburne (52), Glock (62), and Renshaw (142).   None demonstrates conclusively that flashmeter training or eye-movement training is superior to other methods of training in reading.

Renshaw (142) has emphasized the transferability   of   tachistoscopic training, not only for reading but for such activities as plane recognition, on the assumption that it trains a general "perceptual skill" which he compares with motor skills. He reported that Navy preflight trainees improved their skill at plane recognition by tachistoscopic training with digits and slides containing varied numbers of planes to count ("counter training").   But no data were presented for a control group without tachistoscopic training. The experiment was repeated with an appropriate control group by the AAF Psychological Test Film Unit (57). Tests

for recognition proficiency showed no advantage for the group which received tachistoscopic training, although their performance with the digit and counter slides improved. A somewhat similar experiment was conducted in an Air Force Gunnery Training program (76). The Os were given practice in observing, or observing and drawing patterns exposed on a large screen for .05 sec. Pre- and posttests on five criteria gave little evidence of transfer. A test of speed of plane identification resulted in a better score for a control group. It appears likely, in view of all the evidence, that perceptual learning with the tachistoscope transfers only insofar as the test and training tasks are similar.

Lindsley (112, 113) used a "flash-reading trainer" for radar operators with some success. The trainer simulated the presentation of the actual PPI type radar-scope. Signal blips were flashed briefly (2.5 sec.) on a screen marked with range rings and bearing lines or with a rectangular grid. Different problems (e.g., reading the location of a blip or counting blips) were given during practice. Improvement in accuracy of readings increased with training, and correlations showed a significant relationship between trainer proficiency and actual proficiency in operating the PPI scope. In this experiment, the trainer task was made as similar as possible to the operation to be performed.

Other types of transfer of perceptual learning have been investigated with positive results. Coover and Angell (25) thought that general habituation to experimental conditions was responsible for transfer. Their Os were tested for discrimination of shades of gray; then they were given training in discrimination of sound intensities for 17 days, and

finally were retested on discrimination of the grays. Gilbert and Fracker (60) found that discriminative reaction time showed a similar transfer from auditory to visual discrimination.

Transfer of a general principle was demonstrated by Judd's (94) study of dart throwing at an underwater target. Knowledge of the general principle of refraction had a transfer effect when the spatial cues were changed. Hendrickson and Schroeder (72) repeated the experiment with an air gun; they minimized the motor control required, and found both original performance and performance on the changed problem facilitated by knowledge. Minturn and Reese found increased accuracy of estimation for dot patterns as yet unpracticed, following correction with a similar pattern (125). Here O could verbalize a tendency to constant error and apply it throughout the stimulus dimension.

Bilateral transfer and transfer to other areas of a sensory surface have been studied with the two-point limen and with peripheral recognition of form. Practice which lowered the two-point limen showed a similar lowering of the limen on corresponding symmetrical areas of the skin (33, 128, 181). Transfer to adjacent areas of the homolateral side was obtained by Mukherjee (128) but not by Volkmann (181). Franz and his collaborators (48, 49, 50) found that practice which increased accuracy of recognition of forms presented peripherally transferred to the corresponding retinal area of the opposite eye. Homolateral transfer to untrained retinal areas also occurred (50) but to a lesser degree. The areas tested in the latter case were farther out on the periphery than the trained area.

An experiment by Fracker (47)

suggests that transfer may also be mediated by a learned "conceptual schema." His Os were asked to reproduce the order in which four tones, differing in intensity, were presented to them. The tones were labelled 1, 2, 3, and 4, according to intensity, and were presented in different patterns. After long practice, the Os were asked to reproduce orders for other series—four grays, nine tones of different intensity, nine grays, and four tones of different pitch. Highest transfer occurred for the four grays, and different degrees for the others. Introspections indicated that O had learned to use a spatial schema for ordering intensities, which could be carried over to the four grays.

How transfer operates when stimuli are shifted along a continuum is a question of theoretical interest since it is related to the problem of transposition and to the question of whether a response to a given stimulus quantity is originally specific or is only relative. Cameron (18) had Os practice singing one standard tone for several months. Relative discrimination was tested, and this was improved by the practice only when the practiced standard was one of the pair to be discriminated. The outcome suggests that training in absolute identification of a given tone can reduce stimulus generalization between that particular point and neighboring ones but does not increase specificity for other points in the scale. Wedell (187) studied the training of absolute pitch, in order to discover whether O learned to recognize specific individual notes, or learned the position of notes with respect to a whole scale. If the latter occurs, Wedell thought he should obtain some ability to identify notes never practiced, provided they are

within the range of notes practiced. He gave practice with 25 notes equidistant in pitch, which had to be identified by vibration rate. The O had before him a chart representing the notes he was learning, each designated by its vibration rate. In the last two test sessions, either 9 of the 25 or all of the 25 tones were replaced by ones near them, but different in frequency. The new tones did not produce any increase in the magnitude of error (measured in DL's) over the last training session. Wedell concluded that Os developed a subjective scale in which they could place unpracticed tones as well as practiced ones, and that they learned the scale rather than individual notes.

In this interesting experiment the result of O's learning was to increase the specificity of the correspondence between the stimulus input and the perceptual judgment. Average error in DL's was reduced by one-half or more. One might ask how an increased specificity of stimulus-response connections can transfer to other stimuli when the very concept of transfer implies nonspecificity or generality of connections. The answer to this seeming paradox may lie in O's ability to make a conceptual generalization as the differential learning proceeds, which can then mediate transfer when the particular stimuli are shifted along the dimension. Here a subjective scale with anchored and differentiated points provided a common yardstick on which new stimuli could be located. When dimensional transfer fails, the learning situation may be one which does not permit development of a useful conceptual mediator. An experiment by Gibson and Smith (56) may illustrate this point. The Os were given training in estimating

different distances. Each distance was represented by a photograph of a stake driven into the ground in a very long, level field. The *O*s soon learned to give the appropriate distance in yards for the stakes. But when size judgments of the stakes (by the method of matching with a standard) were later asked for, accuracy (that is, size constancy) was not improved by the training in distance estimation. Introspections showed that *O* was not developing a conceptual depth scale during the training at all; instead he was memorizing specific cues, such as accidental spots in the photograph, for the different estimates. Such learning could not transfer to the new judgment of size-at-a-distance.

RETENTION

Facts having to do with retention of perceptual learning are spotty and permit no generalization, except that forgetting over time does occur. Whether it occurs because of sheer disuse or some positive cause is not known, though Postman and Page (138) have demonstrated retroactive inhibition in a task involving perceptual learning. No study has secured sufficient evidence to plot a curve of the extent of forgetting with time. Degree of retention appears to be different depending on the nature of the learned performance. For instance, there is "rapid loss with disuse" of improvement in the two-point limen on the skin (33, 128); but improved visual acuity was said to be retained after as long as two years (120). Improved peripheral acuity was said to show about 50 per cent retention after various intervals (114, 115). Dallenbach reported that the effects of span-of-apprehension training in children persisted 41 weeks or longer after practice ceased (29, 30).

The effect of training in judgment of numerosity (125) was still apparent after six months, though variability rose again. An illusion which had been reduced with practice still showed a CE smaller than the original one after a year (19). Improvement in discrimination of absolute pitch in one study (124) was reported largely lost after several years of no practice. In another study of absolute pitch (63) 16 *O*s retested after a year showed good retention, seemingly proportional to the percentage of improvement during training. Clarification of these results awaits systematic experimentation.

DISCUSSION

Theoretical discussions of perceptual learning have been fairly frequent in recent years, but on the whole they contribute little to the explanation of the facts presented in this paper. Helmholtz's "unconscious inference" (70) is the parent of most current theories. Ames and his followers (20, 86, 96) have a similar concept in the form of implicit "assumptions" which are gained from past experience and mediate perception. A given retinal pattern is said to be perceived as a chair, instead of a mass of lines or a cat's cradle or anything else, because in the past this pattern of stimulation was associated with sitting in it. The stimulation has no prognostic value for perception; instead, the perception itself is prognostic because of the assumptions which mediate it. Perceiving is the "apprehending of probable significances" (86, p. 290).[5]

[5] Cf. Titchener's remark "We may believe with Cattell that 'perceptions are . . . in large measure the result of experience and utility,' but they must still have a psychophysical substrate, on the one hand; and, on the other, the bare reference to utility does not explain them" (166, p. 208).

Yet the evidence for perceptual learning which has been presented in the previous pages requires the conclusion that practice results in a *closer approximation of discriminative responses to differential stimulation.* Lowered DL's for pitch, or improvement in absolute estimation of visual dimensions such as distance, both mean that perceptual judgments have changed in the direction of closer and more specific relationship to the stimulus input. How can this happen if perception is determined by the purpose of the observer and his assumptions (see **96**, pp. 88 ff.), and has only a fortuitous, noncorrelative relationship with stimulation? It seems obvious that there is a dimensional correspondence (cf. **59**, **141**), perhaps very crudely differentiated to begin with; for example, *O* reacts to a variation in tonal frequency with some change in his experience of pitch if the variation is large. With training, his perceptions become better differentiated and permit finer discriminations within the dimension. But how can "assumptions" accomplish this?

Brunswik's functional and "ecological" theory of perception (**14**, **15**) is similar to Ames's theory in assuming that any correspondence between perception and stimulation is learned. The environment provides certain ecological relationships of objects and events which the perceiver happens to experience in context or sequence with varying frequency. Perception of a given object depends on the frequency of association of "cue" and "referent" in past experience. The observer, consciously or not, makes an interpretation in terms of *probability* of referent value of the stimulation, the probability deriving from previous occurrences of this relationship. A computer analogy has been suggested for this process (**75**). This view, that perception has the function of keeping *O* in touch with a distal world by rating the incoming cues as to their validity does not seem to account for the facts of refinement of discrimination within a dimension, at least in its present formulation.

Bruner (**13**) and Postman (**135**) have both expressed a somewhat different formulation of how learning operates in perception. "We can conceive of the perceptual process as a cycle of hypothesis—information—trial and check of hypothesis—confirmation or non-confirmation" (**135**, p. 251). Here, receptive processes are assumed to yield information in themselves; in fact, under some conditions of stimulation they provide "full information" and there is nothing to learn. The theory is designed primarily to apply to ambiguous stimulation—in itself, very difficult to define—and again does not seem to handle increased accuracy of discrimination.

Hebb's empirical theory (**69**) is aimed at the perception of form. Perception is innately organized as figure on ground—the figure has a "primitive unity"—but "identity" of a form involves a developmental process. Perception of a square or circle is slowly learned, he says, and depends originally on multiple visual fixations (pp. 34 ff.). If there are visual fixations, presumably dimensional discrimination of some kind is innate. How further differentiation within dimensions is achieved is not clear.

The theoretical problem might be formulated as follows. If we think of the stimulus variable as spread along a continuum, as it has been in

most of the cases cited here, and the response categories as bearing some relationship which can be fitted along a dimension, lowering of a variable error results in the reduction of generalization curves around points on the stimulus continuum. The solid lines in Fig. 1 (variable error) represent hypothetical generalization curves at the beginning of the experiment; the band of stimulation which may elicit a particular response is originally very wide and the S-R relationship is not very specific. The dotted lines represent hypothetical generalization curves after training; the band of stimulation which may elicit any R is narrower, that is, the S-R relationship has increased in specificity. How might this effect be achieved? A procedure of "scale training," in which stimulation at different points in the scale is responded to with a judgment by O, in turn followed by differential reinforcement (correction), has been effective in many of the experiments cited using the method of single stimuli. The analogy with a conditioning procedure is less direct in experiments where practice is given in relative judgment, though the results are the same. The effectiveness of progressive practice suggests that a quality hitherto not responded to in isolation is being differentiated from the total stimulus input and utilized as a cue variable.

Correction of a constant error is represented in the lower diagram of Fig. 1. It is not hard to account for, since a set to shift the judgments upward or downward may be acquired and can supplement the kind of practice already mentioned.

The above description of the theoretical problem does not apply too well to the cases classified as improved perception under conditions of impoverishment. Here there is generally a distinct pattern to be recognized and identified rather than a continuous dimension of stimulation to be differentiated. In some cases, the learning took the form of paired associates, or what might be called identification learning. An identifying response was learned to "reduced cues," that is, to fragmentary stimulation. In other cases new habits of fixation and scanning may have been important, and in still others, tricks of grouping, etc.

Increased acuity on the skin and retina must be considered in terms of the procedure used for practice. Some cases of increased visual acuity seem to resemble learning with impoverished stimulation, that is, O learns to respond to "reduced cues" by learning what the fragmentary stimuli represent. But in other cases, the process seems to involve learning to respond to a quality of stimulation hitherto undifferentiated. For example, reduction of the two-point limen on the skin was considered by Boring a case of O's learning to use a finer criterion for identification—a better acquaintance with the difference in feel between two points and one (8).

The fact that there may be some transfer of improved discrimination suggests that a process of abstraction may go on with the development of differentiation. At least, O seems to abstract the quality common to the scale, that is, he gets a concept of the dimension of stimulation being varied. He might next, under some conditions of training, conceptualize the scale unit, the ends of the scale, and perhaps proportional intervals. An experiment by Werner on "micro-melodies" (190) suggests such a

REDUCTION OF VARIABLE ERROR
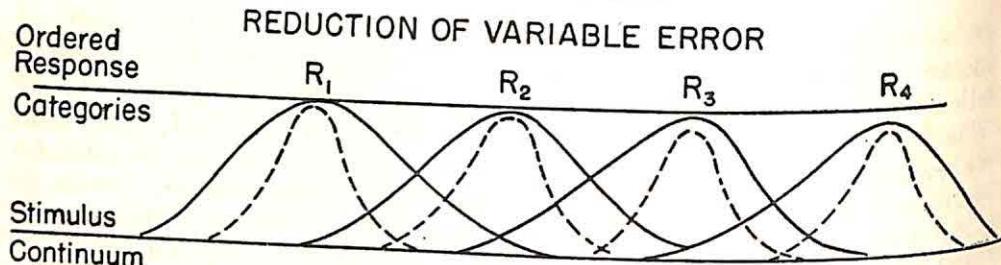
Each solid distribution indicates the range of stimuli which might elicit a given response before training; each broken one indicates the range which might elicit the response after a hypothetical process of differentiation produced by training.

REDUCTION OF CONSTANT ERROR
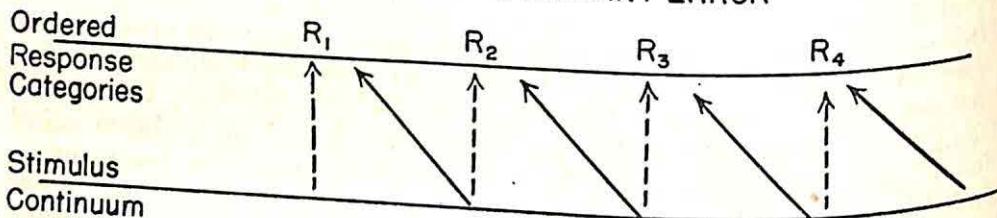
The solid arrows indicate an S—R relationship which might exist before training; the broken arrows, a relationship which might exist after hypothetical correction of a constant error.

FIG. 1. A SCHEMATIC FORMULATION OF CHANGES WHICH MAY OCCUR IN IMPROVEMENT OF PERCEPTUAL JUDGMENTS

process. His Os learned a whole new microscale which had 12 distinguishable notes in a pitch range so narrow that before training the discriminations were impossible. Four Os learned to identify the tones with numbers when a melody was played and were able to recognize a melody transposed one tone up or down. If one tone was altered in the transposition, O could often tell which one it was.

That perceptual learning occurs under many conditions is clear, as also the fact that improved skill in discrimination is an important feature of such learning. Other kinds of perceptual learning also occur, such as learning to recognize and identify objects, a topic too large to have been included within the present review. An enormous array of problems invites investigation in this area and provides a program of research which may yield practical results as well as the eventual enrichment of learning theory.

REFERENCES

1. ALLBAUGH, R., & MILLER, C. Techniques used for improving visual efficiency. *Proc. Iowa Acad. Sci.*, 1946, 53, 263–268.

2. ARNOULT, M. D. Transfer of pre-differentiation training in simple and multiple shape discrimination. *J. exp. Psychol.*, 1953, 45, 401–409.

3. BAKER, KATHERINE E., & WYLIE, RUTH C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, 40, 632–638.

4. BARTLETT, F. C. *Remembering.* Cambridge: Cambridge Univer. Press, 1932.

5. BEVAN, W., & ZENER, K. Some influences of past experience upon the perceptual thresholds of visual form. *Amer. J. Psychol.*, 1952, 65, 434–442.

6. BIEL, W. C., & BROWN, G. E. *Estimation of airplane speed and angle of approach.* Tufts College Project 505-6, NDRC Appl. Psychol. Panel and ONR Project 143-151. Denison Univer., Granville, Ohio, 1949.

7. BLACK, J. W. Final report in summary of work on voice communication. OSRD, 1945; Publ. Bd., No. 12051. Washington: U. S. Dept. of Commerce, 1946.

8. BORING, E. G. The control of attitude in psychophysical experiments. *Psychol. Rev.*, 1920, 27, 440–452.

9. BRIDGMAN, C. S., & HOFSTETTER, H. W. "Improving" color vision. *Optom. Wkly*, 1943, 34, 471–473.

10. BROWN, W. The judgment of differences. *Univer. Calif. Publ. Psychol.*, 1910, 1, 1–71.

11. BROWN, W. The judgment of very weak sensory stimuli. *Univer. Calif. Publ. Psychol.*, 1914, 1, 199–268.

12. BRUCE, R. H., & LOW, F. N. The effect of practice with brief-exposure techniques upon central and peripheral visual acuity and a search for a brief test of peripheral acuity. *J. exp. Psychol.*, 1951, 41, 275–280.

13. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 121–147.

14. BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. of Calif. Press, 1947.

15. BRUNSWIK, E. The conceptual framework of psychology. *Int. Encycl. unif.*

Sci., Vol. I, No. 10. Chicago: Univer. of Chicago Press, 1952.

16. BRUNSWIK, E., & HERMA, H. Probability learning of perceptual cues in the establishment of a weight illusion. *J. exp. Psychol.*, 1951, 41, 281–289.

17. BURRI, CLARA. Process of learning simultaneous binocular vision. *Arch. Ophthal.*, 1942, 28, 235–244.

18. CAMERON, E. H. Effects of practice in the discrimination and singing of tones. *Psychol. Monogr.*, 1917, 23, No. 3 (Whole No. 100), 159–180.

19. CAMERON, E. H., & STEELE, W. M. The Poggendorff illusion. *Psychol. Monogr.*, 1905, 7, No. 1 (Whole No. 29), 83–111.

20. CANTRIL, H., AMES, A., HASTORF, A. H., & ITTELSON, W. H. Psychology and scientific research. *Science*, 1949, 110, 461–464, 491–497, 517–522.

21. CAPURSO, A. A. The effect of an associative technique in teaching pitch and interval discrimination. *J. appl. Psychol.*, 1934, 18, 811–818.

22. CHAPANIS, A. The stability of "improvement" in color vision due to training—a report of three cases. *Amer. J. Optom.*, 1949, 26, 251–259.

23. CHODROFF, M. N. Squint—the psychophysiological aspects involved in its treatment. *Amer. J. Optom.*, 1947, 24, 433–437.

24. CONNETTE, E. The effect of practice with knowledge of results. *J. educ. Psychol.*, 1941, 32, 7, 523–532.

25. COOVER, J. F., & ANGELL, F. General practice effect of special exercise. *Amer. J. Psychol.*, 1907, 18, 328–340.

26. COTZIN, M., & DALLENBACH, K. M. "Facial vision": the role of pitch and loudness in the perception of obstacles by the blind. *Amer. J. Psychol.*, 1950, 63, 485–515.

27. COUEY, F. The Air University Program of reading training through improvement of visual processes. *Amer. Psychologist*, 1948, 3, 294. (Abstract)

28. CROSLAND, H. R., TAYLOR, H. R., & NEWSOM, S. J. Practice and improvability in the Müller-Lyer illusion in relation to intelligence. *J. gen. Psychol.*, 1929, 2, 290–306.

29. DALLENBACH, K. M. The effect of practice upon visual apprehension in school children. Part I & II. *J. educ. Psychol.*, 1914, 5, 321–334, 387–404.

30. DALLENBACH, K. M. The effect of practice upon visual apprehension in the

feeble-minded. *J. educ. Psychol.*, 1919, 10, 61–82.

31. DOBROWOLSKY, W., & GAINE, A. Über die Sehscharfe (Formsinn) an der Peripherie der Netzhaut. *Pflugers Arch. f. d. ges. Physiol.*, 1876, 12, 411 ff.

32. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

33. DRESSLAR, F. B. Studies in the psychology of touch. *Amer. J. Psychol.*, 1894, 6, 313–368.

34. DRURY, M. B. Progressive changes in non-foveal perception of line patterns. *Amer. J. Psychol.*, 1933, 45, 628–646.

35. DVORINE, I. Reconditioning the color-blind; a case report. *Amer. J. Optom.*, 1944, 21, 508–510.

36. DVORINE, I. Improvement in color vision in 20 cases. *Amer. J. Optom.*, 1946, 23, 302–321.

37. EAGLESON, O. W. Comparative studies of white and Negro subjects learning to discriminate visual magnitude. *J. Psychol.*, 1937, 4, 167–197.

38. EGAN, J. P. Articulation testing methods. *Laryngoscope*, 1948, 58, 955–991.

39. ELLSON, D. G. Hallucinations produced by sensory conditioning. *J. exp. Psychol.*, 1941, 28, 1–20.

40. EVANS, R. N. Training improves micrometer accuracy. *Personnel Psychol.*, 1951, 4, 231–242.

41. FEHRER, Elizabeth V. An investigation of the learning of visually perceived forms. *Amer. J. Psychol.*, 1935, 47, 187–221.

42. FERNBERGER, S. W. The effect of practice in its initial stages in lifted weight experiments and its bearing upon anthropomorphic measurements. *Amer. J. Psychol.*, 1916, 27, 261–272.

43. FERNBERGER, S. W. On absolute and relative judgments in lifted-weight experiments. *Amer. J. Psychol.*, 1931, 43, 560–578.

44. FERNBERGER, S. W., & IRWIN, F. W. Time relations for the different categories of judgments in the "absolute method" in psychophysics. *Amer. J. Psychol.*, 1932, 44, 505–525.

45. FISCHOFF, S. C. A report on reconditioning of color blindness. *Optom. Wkly*, 1943, 34, 960–961.

46. FOSTER, W. S. The effect of practice upon visualizing and upon the reproduction of visual impressions. *J. educ. Psychol.*, 1911, 2, 11–22.

47. FRACKER, G. C. On the transference of training in memory. *Psychol. Monogr.*, 1908, 9, No. 2 (Whole No. 38), 56–102.

48. FRANZ, S. I., & LAYMAN, J. D. Studies in cerebral function. I. Peripheral retinal learning and practice transfer. *Publ. Univer. Calif. Los Angeles, Educ., Phil., Psychol.*, 1933, 1, 65–78.

49. FRANZ, S. I., & KILDUFF, S. Studies in cerebral function. II. Cerebral dominance as shown by segmental visual learning. *Publ. Univer. Calif. Los Angeles, Educ., Phil., Psychol.*, 1933, 1, 79–90.

50. FRANZ, S. I., & MORGAN, R. C. Studies in cerebral function. III. Transfer of effects of learning from one retinal area to other retinal areas. *Publ. Univer. Calif. Los Angeles, Educ., Phil., Psychol.*, 1933, 1, 91–98.

51. FRANZ, S. I., & DAVIS, E. F. Studies in cerebral function. IV. Simultaneous reading with both cerebral hemispheres. *Publ. Univer. Calif. Los Angeles, Educ., Phil., Psychol.*, 1933, 1, 99–105.

52. FREEBURNE, C. M. The influence of training in perceptual span and perceptual speed upon reading ability. *J. educ. Psychol.*, 1949, 40, 321–352.

53. GAGNÉ, R. M., & BAKER, KATHERINE E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, 40, 439–451.

54. GALLAGHER, J. R., LUDVIGH, E. J., MARTIN, S. F., & GALLAGHER, C. D. Effect of training methods on color vision. *Arch. Opthal.*, 1947, 37, 572–582.

55. GATES, A. I. Functions of flash-card exercises in reading: an experimental study. *Teachers Coll. Rec.*, 1925, 27, 311–327.

56. GIBSON, ELEANOR J., & SMITH, JoANN. The effect of training in distance estimation on the judgment of size-at-a-distance. Human Resources Research Center, Lackland Air Force Base, *Research Bulletin*, 1952, No. 52–39.

57. GIBSON, J. J. (Ed.) *Motion picture testing and research.* Report No. 7, AAF Aviation Psychology Program Research Reports. Washington: U. S. Govt. Print. Off., 1947.

58. GIBSON, J. J. Studying perceptual phenomena. In T. G. Andrews (Ed.), *Methods of psychology.* New York: Wiley, 1948. Pp. 158–188.

59. GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.

60. GILBERT, J. A., & FRACKER, G. C. The

effects of practice in reaction and discrimination for sound upon the time of reaction and discrimination for other forms of stimuli. *Univer. Iowa Stud. Psychol.*, 1897, 1, 62–76.

61. GILLILAND, A. R. The effect of practice with and without knowledge of results in grading handwriting. *J. educ. Psychol.*, 1925, 16, 532–536.

62. GLOCK, M. D. The effect upon eye-movements and reading rate at the college level of 3 methods of training. *J. educ. Psychol.*, 1949, 40, 93–106.

63. GOUGH, E. The effects of practice on judgments of absolute pitch. *Arch. Psychol.*, 1922, 7, No. 47.

64. GRETHER, W. F., & WOLFLE, D. L. The relative efficiency of constant and varied stimulation during learning. II. White rats on a brightness discrimination problem. *J. comp. Psychol.*, 1936, 22, 365–374.

65. GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.

66. HALSEY, RITA M., & CHAPANIS, A. On the number of absolutely identifiable spectral hues. *J. opt. Soc. Amer.*, 1951, 41, 1057–1058.

67. HAMILTON, H. C. The effect of incentives on accuracy of discrimination measured on the Galton bar. *Arch. Psychol.*, 1929, 16, No. 103.

68. HARRIMAN, A. E., & MACLEOD, R. B. Salt discrimination thresholds in normal and adrenalectomized rats under conditions of thirst motivation and electric shock punishment. *Amer. J. Psychol.*, 1953, 66, 465–471.

69. HEBB, D. O. *The organization of behavior.* New York: Wiley, 1949.

70. HELMHOLTZ, H. VON. *Physiological optics.* (Trans. by J. P. C. Southall.) Vol. III. Opt. Soc. Amer., 1924.

71. HELSON, H. Adaptation level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297–313.

72. HENDRICKSON, G., & SCHROEDER, W. H. Transfer of training in learning to hit a submerged target. *J. educ. Psychol.*, 1941, 32, 205–213.

73. HENLE, MARY. An experimental investigation of past experience as a determinant of visual form perception. *J. exp. Psychol.*, 1942, 30, 1–21.

74. HENRY, L. K., & LAUER, A. R. A comparison of four methods of increasing the reading speed of college students. *Proc. Iowa Acad. Sci.*, 1939, 46, 273–276.

75. HILGARD, E. R. The role of learning in perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 95–120.

76. HOBBS, N. (Ed.) *Psychological research on flexible gunnery training.* Report No. 11, AAF, Aviation Psychology Program Research Reports. Washington: U. S. Govt. Print. Off., 1947.

77. HOISINGTON, L. B. An example of the fractionation of data from the method of constant stimuli for the two-point limen. *Amer. J. Psychol.*, 1917, 28, 588–596.

78. HOLLINGWORTH, H. L. The inaccuracy of movement. *Arch. Psychol.*, 1909, No. 13.

79. HOLWAY, A. H., JAMESON, DOROTHEA A., ZIGLER, M. J., HURVICH, L. M., WARREN, A. B., & COOK, E. B. *Factors influencing the magnitude of range errors in free space and telescopic vision.* Boston: Division of Research, Graduate School of Business Administration, Harvard Univer., 1945.

80. HOROWITZ, M. W., & KAPPAUF, W. E. *Aerial target range estimation.* OSRD, Report No. 5301, 1945; Publ. Bd., No. 15812. Washington: U. S. Dept. of Commerce, 1946.

81. HOWELL, W. H. *A text-book of physiology.* (11th Ed.) Philadelphia: W. B. Saunders, 1931.

82. HOWES, D. The intelligibility of spoken messages. *Amer. J. Psychol.*, 1952, 65, 460–465.

83. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, 41, 401–410.

84. HUMES, J. F. The effect of practice upon the upper limen for tonal discrimination. *Amer. J. Psychol.*, 1930, 42, 1–16.

85. IRWIN, F. W. Thresholds for the perception of difference in facial expression and its elements. *Amer. J. Psychol.*, 1932, 44, 1–17.

86. ITTELSON, W. H. The constancies in perceptual theory. *Psychol. Rev.*, 1951, 58, 285–294.

87. JAMES, W. *Principles of psychology.* Vol. I. New York: Henry Holt, 1890.

88. JOHNSON, D. M. Generalization of a scale of values by the averaging of practice effects. *J. exp. Psychol.*, 1944, 34, 425–436.

89. JOHNSON, D. M. A systematic treatment of judgment. *Psychol. Bull.*, 1945, 42, 193–224.

90. JOHNSON, D. M. Generalization of a

reference scale for judging pitch. *J. exp. Psychol.*, 1949, 39, 316–321.

91. JOHNSON, D. M. Learning function for a change in the scale of judgment. *J. exp. Psychol.*, 1949, 39, 851–860.

92. JUDD, C. H. Practice and its effects on the perception of illusions. *Psychol. Rev.*, 1902, 9, 27–39.

93. JUDD, C. H. The Müller-Lyer illusion. *Psychol. Monogr.*, 1905, 7, No. 1 (Whole No. 29), 55–81.

94. JUDD, C. H. The relation of special training to general intelligence. *Educ. Rev.*, 1908, 36, 28–42.

95. JUDD, C. H., & COURTEN, H. C. The Zöllner illusion. *Psychol. Monogr.*, 1905, 7, No. 1 (Whole No. 29), 112–139.

96. KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view.* Office of Naval Research, Contract Nonr-496(01). Hanover, N. H.: Institute for Associated Research, 1952.

97. KOFFKA, K. *Growth of the mind.* New York: Harcourt, Brace, 1924.

98. KÖHLER, W., & FISHBACK, JULIA. The destruction of the Müller-Lyer illusion in repeated trials: I. An examination of two theories. *J. exp. Psychol.*, 1950, 40, 267–281.

99. KÖHLER, W., & FISHBACK, JULIA. The destruction of the Müller-Lyer illusion in repeated trials: II. Satiation patterns and memory traces. *J. exp. Psychol.*, 1950, 40, 398–410.

100. LAMBERT, W. W., SOLOMON, R. L., & WATSON, P. D. Reinforcement and extinction as factors in size estimation. *J. exp. Psychol.*, 1949, 39, 637–641.

101. LANCASTER, W. B. Present status of eye exercises for improvement of visual function. *Arch. Opthal.*, 1944, 32, 167–172.

102. LANGE, C. W. Modern visual development. *Optom. J. Rev. Optom.*, 1947, 83, 37–40.

103. LASHLEY, K. S. The mechanism of vision: XV. Preliminary studies of the rat's capacity for detail vision. *J. genet. Psychol.*, 1938, 18, 123–193.

104. LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, 39, 770–784.

105. LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, 40, 175–188.

106. LAWRENCE, D. H. Generalization gradients and the transfer of a discrimina-

tion along a continuum. *J. comp. physiol. Psychol.*, 1952, 45, 511–516.

107. LEEPER, R. A study of a neglected portion of the field of learning—the development of sensory organization. *J. genet. Psychol.*, 1935, 46, 41–75.

108. LEHMANN, A. Ueber Wiedererkennen. *Philos. Stud. (Wundt)*, 1889, 5, 96–156.

109. LEPPER, J. H. More about reconditioning the color blind. *Opt. J.*, 1942, 79, 20.

110. LEWIS, E. O. The effect of practice on the perception of the Müller-Lyer illusion. *Brit. J. Psychol.*, 1908, 2, 294–306.

111. LICKLIDER, J. C. R., & POLLACK, I. Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech. *J. acoust. Soc. Amer.*, 1948, 20, 42–51.

112. LINDSLEY, D. B., *et al.* *A radar trainer and flash reading method for operators of the Plan Position Indicator.* OSRD, 1944; Publ. Bd., No. 18334. Washington: U. S. Dept. of Commerce, 1946.

113. LINDSLEY, D. B., *et al.* *Use of the P.P.I. flash-reading trainer in training Navy search-radar operators.* OSRD, 1945; Publ. Bd., No. 18336. Washington: U. S. Dept. of Commerce, 1946.

114. LOW, F. N. *Effect of training on acuity of peripheral vision.* Civil Aeronautics Administration, Div. Research, Report No. 68. Washington: 1946.

115. LOW, F. N. Some characteristics of peripheral visual performance. *Amer. J. Physiol.*, 1946, 146, 573–584.

116. LUCIANI, L. *Human physiology.* Vol. IV. (Trans. by F. A. Welby.) London: Macmillan, 1917.

117. MacLATCHY, J. Bexley reading study. *Educ. Res. Bull.*, Ohio State Univer., 1946, 25, 141–168.

118. MASON, H. M., & GARRISON, B. K. Intelligibility of spoken messages; liked and disliked. *J. abnorm. soc. Psychol.*, 1951, 46, 100–103.

119. McFADDEN, H. B. *Visual acuity.* Duncan, Okla.: Optometric Extension Program, 1940.

120. McFADDEN, H. B. *Three studies in psychological optics. I. The permanence of the effects of training on visual acuity.* Duncan, Okla.: Optometric Extension Program, 1941. Pp. 1–10.

121. McFADDEN, H. B. *Three studies in psychological optics. II. A revised method for studying the effects of training on visual acuity.* Duncan, Okla.: Optometric Extension Program, 1941,

Pp. 11–26.

122. McGarvey, Hulda R. Anchoring effects in the absolute judgment of verbal materials. *Arch. Psychol.*, 1943, No. 281.

123. McGinnies, E., Comer, P. B., & Lacey, O. L. Visual recognition thresholds as a function of word length and word frequency. *J. exp. Psychol.*, 1952, 44, 65–69.

124. Meyer, M. Is the memory of absolute pitch capable of development by training? *Psychol. Rev.*, 1899, 6, 514–516.

125. Minturn, A. Leigh, & Reese, T. W. The effect of differential reinforcement on the discrimination of visual number. *J. Psychol.*, 1951, 31, 201–231.

126. Moers, M. Ein Beitrag zur Untersuchung der Augenmassprüfung. *Z. f. angew. Psychol.*, 1924, 23, 257–292.

127. Morgan, M. W. An investigation of the use of stereoscopic targets in orthoptics. *Amer. J. Optom.*, 1947, 24, 411–432.

128. Mukherjee, K. C. The duration of cutaneous sensation (I), and the improvement of its sensible discrimination by practice (II). *J. exp. Psychol.*, 1933, 16, 339–342.

129. Mull, H. K. The acquisition of absolute pitch. *Amer. J. Psychol.*, 1925, 36, 469–493.

130. Nash, Myrtle C. A quantitative study of effects of past experience on adaptation level. Unpublished doctor's dissertation, Bryn Mawr College. Ann Arbor, Mich.: University Microfilms, Publ. No. 3632, 1949.

131. Nichols, A. S. *Primary procedures in vision training and the pointer method of vision training with the "AN" series.* Meadville, Pa.: Keystone View Co., 1946.

132. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.

133. Peppard, H. M. *Sight without glasses.* New York: Permabooks, 1948.

134. Petran, L. A. An experimental study of pitch recognition. *Psychol. Monogr.*, 1932, 42, No. 6 (Whole No. 193).

135. Postman, L. Toward a general theory of cognition. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 242–272.

136. Postman, L., & Bruner, J. S. Hypothesis and the principle of closure: the effect of frequency and recency. *J. Psychol.*, 1952, 33, 113–125.

137. Postman, L., Bruner, J. S., & Walk, R. D. The perception of error. *Brit. J. Psychol.*, 1951, 42, 1–10.

138. Postman, L., & Page, R. Retroactive inhibition and psychophysical judgment. *Amer. J. Psychol.*, 1947, 60, 367–377.

139. Postman, L., & Schneider, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, 58, 271–284.

140. Postman, L., & Solomon, R. L. Perceptual sensitivity to completed and incompleted tasks. *J. Pers.*, 1950, 18, 347–357.

141. Pratt, C. C. The role of past experience in visual perception. *J. Psychol.*, 1950, 30, 85–107.

142. Renshaw, S. The visual perception and reproduction of forms by tachistoscopic methods. *J. Psychol.*, 1945, 20, 217–232.

143. Ritter, Sarah M. The vertical-horizontal illusion. *Psychol. Monogr.*, 1917, 23, No. 4 (Whole No. 101).

144. Rogers, M. H., Sprol, S. J., Viteles, M. S., Voss, H. A., & Wickens, D. D. *Evaluation of methods of training in estimating a fixed opening range.* OSRD Report No. 5765, 1945; Publ. Bd., No. 4021. Washington: U. S. Dept. of Commerce, 1946.

145. Rogers, S. The anchoring of absolute judgments. *Arch. Psychol.*, 1941, No. 261.

146. Rossman, Irma L., & Goss, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, 42, 173–182.

147. Saltzman, I. J., & Garner, W. R. Reaction-time as a measure of span of attention. *J. Psychol.*, 1948, 25, 227–241.

148. Sanford, E. C. The relative legibility of the small letters. *Amer. J. Psychol.*, 1888, 1, 402–435.

149. Scott, R. E. Flash cards as a method of improving silent reading in the third grade. *J. educ. Method.*, 1925, 5, 102–112.

150. Seashore, C. E., Carter, E. A., Farnum, Eva C., & Sies, R. W. The effect of practice on normal illusions. *Psychol. Monogr.*, 1908, 9, No. 2 (Whole No. 38), 103–148.

151. Seashore, H., & Bavelas, A. The functioning of knowledge of results in Thorndike's line-drawing experiment. *Psychol. Rev.*, 1941, 48, 155–164.

152. Seashore, R. H. Work methods: an often neglected factor underlying individual differences. *Psychol. Rev.*, 1939, 46, 123–141.

153. Seward, J. P. The effect of practice on the visual perception of form. *Arch. Psychol.*, 1931, 20, No. 130.

154. Smith, F. O. The effect of training in pitch discrimination. *Psychol. Monogr.*, 1914, 16, No. 3 (Whole No. 69), 67–103.

155. Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition threshold for words. *J. exp. Psychol.*, 1952, 43, 195–201.

156. Solomons, L. Discrimination in cutaneous sensations. *Psychol. Rev.*, 1897, 4, 246–250.

157. Stanton, Hazel M., & Koerth, Wilhelmine. Musical capacity measures of adults repeated after musical education. *Univer. Iowa Stud. Ser. on Aims and Prog. of Res.*, 1930, No. 31.

158. Stanton, Hazel M., & Koerth, Wilhelmine. Musical capacity measures of children repeated after musical training. *Univer. Iowa Stud. Ser. on Aims and Prog. of Res.*, 1933, No. 42.

159. Supa, M., Cotzin, M., & Dallenbach, K. M. "Facial vision": the perception of obstacles by the blind. *Amer. J. Psychol.*, 1944, 57, 133–183.

160. Sutherland, J. The relation between perceptual span and rate of reading. *J. educ. Psychol.*, 1946, 37, 373–380.

161. Taubman, R. E. The effect of practice with and without reinforcement on the judgment of auditory number. *J. exp. Psychol.*, 1944, 34, 143–151.

162. Tawney, G. Ueber die Wahrnehmung zweier Punkte mittelst des Tastsinnes, mit Rücksicht auf die Frage der Uebung. *Phil. Stud.*, 1897, 13, 163–222.

163. Thorndike, E. L. *The fundamentals of learning.* New York: Teachers Coll., Columbia Univer., Bureau of Publications, 1932.

164. Thorndike, E. L., & Woodworth, R. S. The influence of improvement in one mental function upon the efficiency of other functions. (I): II. The estimation of magnitudes: III. Functions involving attention, observation and discrimination. *Psychol. Rev.*, 1901, 8, 247–261, 384–395, 553–564.

165. Tinker, M. A. Visual apprehension and perception in reading. *Psychol. Bull.*, 1929, 26, 223–240.

166. Titchener, E. B. *Experimental Psychology. Vol. I. Qualitative experiments. Part II. Instructor's manual.* New York: Macmillan, 1901.

167. Titchener, E. B. *Experimental Psychology. Vol. II. Quantitative experiments. Part II. Instructor's manual.* New York: Macmillan, 1905.

168. Tresselt, M. E. The influence of amount of practice upon the formation of a scale of judgment. *J. exp. Psychol.*, 1947, 37, 251–260.

169. Trowbridge, M. H., & Cason, H. An experimental study of Thorndike's theory of learning. *J. gen. Psychol.*, 1932, 7, 245–258.

170. Ueno, Y. Experiments on perceptual judgment of space bisection. *Jap. J. Psychol.*, 1926, 1, 453–475.

171. Urban, F. M. Der Einfluss der Uebung bei Gewichtsversuchen. *Arch. f. d. ges. Psychol.*, 1913, 29, 271–311.

172. Van de Water, Marjorie. Seeing in the dark; Navy lookouts are given special training with ship models displayed on dimly lighted stage to enable them to get the first look at the enemy. *Sci. News Lett.*, 1943, 43, 314–316.

173. Van Tuyl, M. C. Monocular perception of distance. *Amer. J. Psychol.*, 1937, 49, 515–542.

174. Van Voorhis, W. R. The improvement of space perception ability by training. Unpublished doctor's dissertation, Penn. State Coll., 1941.

175. Vernon, Margaret D. *The experimental study of reading.* London: Cambridge Univer. Press, 1931.

176. Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, 44, 273–282.

177. Verplanck, W. S., *et al. Response mechanisms at the visual threshold; a methodological study.* Status Report II, Project NR 140–015, Harvard Univer., October, 1952.

178. Viteles, M. S., *et al. An investigation of the range estimation Trainer, Device 5C-4, as a method of teaching range estimation.* OSRD Report No. 4263, 1944; Publ. Bd., No. 4024. Washington: U. S. Dept. of Commerce, 1946.

179. Viteles, M. S., Gorsuch, J. H., Bayroff, A. G., Rogers, M. H., & Wickens, D. D. *Learning range estimation on the firing line.* OSRD Report No. 4405, 1944; Publ. Bd., No. 4023. Washington: U. S. Dept. of Commerce, 1946.

180. Viteles, M. S., Gorsuch, J. H., & Wickens, D. D. *History and final report of Project N-105, Applied Psychol-*

ogy Panel. OSRD Report No. 6266, 1945; Publ. Bd., No. 4018. Washington: U. S. Dept. of Commerce, 1946.

181. VOLKMANN, A. W. Ueber den Einfluss der Uebung. *Leipzig Berichte, Math.-phys. Classe,* 1858, **10**, 38–69.

182. VOLKMANN, A. W. *Physiologische Untersuchungen im Gebiete der Optik.* Leipzig: Breitkopf und Hartel, 1863.

183. VOLKMANN, J. The method of single stimuli. *Amer. J. Psychol.,* 1932, **44**, 808–809.

184. VOLKMANN, J. The anchoring of absolute scales. *Psychol. Bull.,* 1936, **33**, 742–743. (Abstract)

185. VOSS, H. A., & WICKENS, D. D. *A comparison of free and stadiametric estimation of opening range.* OSRD Report No. 6114, 1945; Publ. Bd., No. 4019. Washington: U. S. Dept. of Commerce, 1946.

186. WEBER, C. O. Effects of practice on the perceptual span for letters. *J. gen. Psychol.,* 1942, **26**, 347–351.

187. WEDELL, C. H. The nature of the absolute judgment of pitch. *J. exp. Psychol.,* 1934, **17**, 485–503.

188. WEDELL, C. H. *Final report in summary of work on the selection and training of night lookouts.* OSRD Report No. 4342, 1944; Publ. Bd., No. 15811. Washington: U. S. Dept. of Commerce, 1946.

189. WELCH, L. The development of discrimination of form and area. *J. Psychol.,* 1939, **7**, 37–54.

190. WERNER, H. Musical "micro-scales" and "micro-melodies." *J. Psychol.,* 1940, **10**, 149–156.

191. WEVER, E. G., & ZENER, K. E. The method of absolute judgment in psychophysics. *Psychol. Rev.,* 1928, **35**, 466–493.

192. WHIPPLE, G. M. Studies in pitch discrimination. *Amer. J. Psychol.,* 1903, **14**, 289–309.

193. WHIPPLE, G. M. The effect of practice upon the range of visual attention and visual apprehension. *J. educ. Psychol.,* 1910, **1**, 249–262.

194. WILCOX, W. W. An interpretation of the relation between visual acuity and light intensity. *J. gen. Psychol.,* 1936, **15**, 405–435.

195. WILLIAMS, M. C. Normal illusions in representative geometrical forms. *Univer. Iowa Stud. Psychol.,* 1902, **3**, 38–139.

196. WOLFE, H. K. On the estimation of the middle of lines. *Amer. J. Psychol.,* 1923, **34**, 313–358.

197. WOLNER, M., & PYLE, W. H. An experiment in individual training of pitch-deficient children. *J. educ. Psychol.,* 1933, **24**, 602–608.

198. WOOD, R. W. *Physical optics.* New York: Macmillan, 1936.

199. WOODROW, H. The relation between abilities and improvement with practice. *J. educ. Psychol.,* 1938, **29**, 215–230.

200. WOODS, A. C. Report from the Wilmer Institute on the results obtained in the treatment of myopia by visual training. *Amer. J. Ophthal.,* 1946, **29**, 28–57.

201. WOODWORTH, R. S. *Experimental psychology.* New York: Henry Holt, 1938.

202. WOODWORTH, R. S. Reenforcement of perception. *Amer. J. Psychol.,* 1947, **60**, 119–124.

203. WORCHEL, P., & DALLENBACH, K. M. "Facial vision": perception of obstacles by the deaf-blind. *Amer. J. Psychol.,* 1947, **60**, 502–553.

204. WRIGHT, FRANCES A. The relation of music endowment and achievement tests to teacher selection. *Yearb. Mus. Educator's Nat. Conf.,* 1928. (Proc. Mus. Supervisor's Nat. Conf. 1928.)

205. WRIGHT, FRANCES A. The correlation between achievement and capacity in music. *J. educ. Psychol.,* 1928, **17**, 50–56.

206. WYATT, RUTH F. Improvability of pitch discrimination. *Psychol. Monogr.,* 1945, **58**, No. 2 (Whole No. 267).

207. YACORZYNSKI, G. K. The threshold of flicker fusion as a function of excitation and inhibition due to conditioning. *J. exp. Psychol.,* 1944, **34**, 335–342.

208. *Errors in visual estimation of range between 800 and 5000 yards of the learning curve with practice and training.* AORG Report No. 140, 1943.

209. HARVARD UNIVERSITY, HOWE LABORATORY OF OPHTHALMOLOGY. *An investigation of the amount of practice advisable in ranging on a simulated-diving-aeroplane target.* NDRC, Project No. 10, Section D-2, 1942; Publ. Bd., No. 55793. Washington: U. S. Dept. of Commerce, 1947.

210. PRINCETON BRANCH, FIRE CONTROL DIVISION. Analysis of range estimation data. Frankford Arsenal: Branch Memorandum No. 20, 1943.

211. *Range finding by subtense methods.* TAR Report No. 8, 1944.

STUDIES OF DREAMING

GLENN V. RAMSEY
Austin, Texas

Reviews of the literature on dreaming have been made in the *Psychological Bulletin* by Frost (35, 36) in 1915 and 1918, by Small (101) in 1920, and by Richards (89) in 1924. The reviewers included articles which were philosophical, literary, and anecdotal as well as those which were more scientific in nature. The present review is restricted to the more empirical and systematic studies on dreaming, particularly those in which some quantitative findings are reported. In general, a study is included in this review if the results presented could be checked by repetition of the investigation. This criterion for selection of articles rules out most but not all of the articles which deal with the analysis and interpretation of dreams. The American literature was rather carefully checked for data on dreaming. Undoubtedly a few reports were missed because they are occasionally buried in investigations of other topics.

A critical evaluation of many of the studies reporting quantitative data on dreaming is often difficult or impossible to make because they are so poorly reported. Some studies report results which appear to have real merit; for others the results appear questionable. The reviewer therefore adopted the policy of presenting findings whenever they were based on quantitative data even though an adequate basis for their evaluation was not included in the respective reports. Conflicting data in the various studies bring into question the validity of certain findings. In addition, future research will continue to act as a refining process in separating the verifiable data from that which is not. For the benefit of future research workers the reviewer discusses at the end of this article a few of the weaknesses he has observed in the published studies and makes a few suggestions for further investigations of the phenomenon of dreaming.

The restriction of the review to the more quantitative studies does not imply that less systematically controlled and reported observations are without value. Many insightful observations have been made in clinical practice which serve as useful hypotheses for experimental and quantitative investigations. The reviewer appreciates the clinical validity of the dynamic nature of dreaming and its relationship to the total functioning of the personality. The complex role and meaning of dreams in the life of the individual are not easily attacked at the present time by quantitative research methods. A few studies of dynamic aspects of dreaming have appeared, and more are certain to follow.

The relatively small number of major research studies on dreams is surprising since dreaming is such a universally experienced phenomenon and has through the centuries been of major interest to man. The lack of scientific interest and data on the topic has contributed to the myriads of speculations and the plethora of unscientific publications in the field. The topic of dreaming is largely shunned by current research workers and is given only the briefest of treatment or entirely omitted in most contemporary psychological textbooks.

Yet this reviewer is of the opinion that research in this area which utilizes the current knowledge of research designs and techniques could yield data which would give a better understanding of dreams. This review is therefore designed to facilitate more systematic research on dreaming by bringing together from diverse sources material bearing upon the topic and by providing a summary of the major findings which are reported therein.

IMAGERY IN DREAMS

Direct inspection of reported dream material has yielded fairly consistent data regarding the nature of imagery appearing in dreams. Studies of imagery in dreams have been made by Calkins (13), Weed and Hallam (115), Manacéine (73), Monroe (80, 81, 82), Andrews (2), Wiggam (118), Thomson (106), Bentley (7), Schriever (95), and Middleton (77). In studies of the dream reports obtained from normal individuals there is general agreement that visual imagery is the predominant type among all the sense modalities that appear in dreams. The second most frequent type of imagery reported is auditory. Visual and auditory imagery together are the principal elements in a large majority of all dreams. Only infrequently do tactual, motor, gustatory, and olfactory imagery occur in dream reports. For a time there was a debate on whether gustatory imagery ever appeared in dreams, but it was shown that this type of imagery did occur in case reports obtained by Titchener (107) and Weed and Hallam (115). A fairly representative set of percentages for the frequency of appearance of various types of imagery in dreams is given by Weed and Hallam (115). They found that out of the 381 dreams reported by six subjects, 84 per cent involved visual imagery, 67 per cent auditory, 10 per cent dermal, 6 per cent gustatory, and 6 per cent olfactory.

The reported figures for the frequency of occurrence of sense modalities in dreams are somewhat difficult to compare because of different methods of calculation. In some cases only the predominant imagery of each dream was tabulated. In other studies all types of imagery appearing in a dream were counted. Sometimes it was impossible to tell which of these two systems was used. In addition a few studies reported frequency figures for the various combinations of imagery which appeared in dreams. Even though direct comparisons cannot be made between many of the studies, the general findings, however, do appear consistent.

Considerable variation among individuals has been noted for the frequency of the various types of imagery appearing in dreams. This point was covered by Weed and Hallam (115), Monroe (82), Thomson (106), and later investigators. Also some variation has been noted in the frequency reports on various types of imagery given by the same individual from time to time.

Visual imagery. Most visual dreams are experienced as being in black and white. This is an interesting deviation from the daytime visual experiences of most individuals. Bentley (7) found that color in dreams appeared in only 20 per cent of the 54 dreams reported by five adult subjects. Among 277 college men and women, Middleton (77) found that 60 per cent had experienced color in dreams. Among individuals who do have color in their dreams, the frequency of such dreams is not high. Monroe (80) reported an

apparent increase in the frequency of colored dreams when his 14 subjects viewed geometric forms and simple objects cut from colored pieces of paper before going to sleep. Morphine has been reported by Jelliffe (55) to cause an increase in the number of colored dreams. Finley (30) claimed that a pituitary hormone given to a patient increased the frequency of colored dreams. Lovett (68) found that psychiatric patients reported more color dreams than did his normal controls. Ellis (27) cites evidence presented by Wynaendts-Franklin that among 300 adults more women than men reported having dreamed in color.

Auditory dreams. Manacéine (73), Calkins (13), and Bentley (7) stated that auditory dreams are primarily composed of verbal sounds and only infrequently were other auditory sounds such as bells, noises, etc. reported. Manacéine (73), however, stated that words, phrases, or sentences appeared in only 8.5 per cent of the dreams he studied. Most frequently the verbal sounds were words spoken by the dreamer, and only infrequently by the other participants in the dream. Some discussion has been made of exclusively auditory dreams which are reported to occur among musicians.

Gustatory. Several incidental reports appear on the effect of various gustatory stimuli upon dreams. For example, Monroe (81) had 22 female students place a clove on their tongues before going to sleep. In the series of dreams reported by them there appeared an increase in the frequency of gustatory elements along with an increase in olfactory elements.

Imagery of the blind. In 1838 Heermann (47) studied the dream imagery of the blind, and his findings were confirmed and extended by later investigators such as Jastrow (54), Wheeler (117), Deutsch (24), and Bolli (12). Jastrow's study in 1888 of 200 blind males and females still remains the basic investigation of this subject. Findings are in agreement that practically all the blind do dream, but possibly not quite so frequently as normals do. The dreams of the congenitally blind were found to be totally lacking in visual imagery, but they were highly filled with auditory and motor images. Dreams of the blind also included tactual, gustatory, organic, and other sense imagery in probably a higher frequency than found in normals. In a case cited by Wheeler (117) a blind individual reported the same type of synesthetic sensations in dreams as experienced in waking life.

Jastrow (54) and others also investigated dream imagery in individuals who became blind at various periods after birth. They reported that individuals who became blind after the first five or six years of life did maintain some visual imagery in their dreams, although they showed a progressive deterioration or fading as the individual became older. A few individuals studied by Jastrow reported visual imagery after 40 years of blindness. Those who became blind before the age of five or six did not retain visual imagery in their dreams.

Imagery of deaf and blind. Jastrow (54), in collaboration with Hall, reported on the dreams of a subject who was both blind and deaf from birth. Her dreams were devoid of both visual and auditory imagery. Tactual and motor imagery were the principal elements in her dreams. The people who entered her dreams "talked" with their fingers and she was observed to be "talking" with her

fingers during her sleep. The authors state, "In short, her dreams are accurately modeled upon experiences of her waking life, reproducing in detail all peculiarities of thought and action which a phenomenal education impressed upon her mind.".

Imagery of the crippled. Jastrow (54) reported on a few cases of imagery in the dreams of the crippled. Those who lost an arm or leg sometime after birth continued to dream of the use of the lost member. For one person, such a dream continued for over 15 years. He cites Heermage as reporting a case of an individual born with only stubs for legs, and in his dreams he walked around on his knees just as he did in actual life.

SPEED OF DREAMS

Many people have speculated on the speed and length of dreams, but as yet no conclusive evidence regarding the topic is at hand. The older writers contended that dreams were exceedingly fast, rarely lasting more than 1 or 2 seconds. Maury's (74) well-known dream report has been frequently cited as evidence of the speed of dreaming. Many case reports and discussions of the speed of dreams have been made since then. Ellis (27) reported the following early studies on the speed of dreaming: Egger (25), Clavière (16), Tobolowska (108), Piéron (86), and Foucault (32). In general, they reported that imagery in dreams was no more rapid than waking imagery, though the illusion of time is experienced by the dreamer in his recall and description. Woodworth (121) attempted to throw light on this topic by timing the speed of hypnogogic reveries and concluded they could be as fast as those reported for dream life.

Others estimate the length of

dreams as much longer, ranging from 1 to 10 minutes. Hacker (42), from self-observation during 450 nights, concluded that dreams last up to 10 minutes. Max (75) found that action currents in the fingers of deaf subjects which were associated with dreaming lasted for as long as $2\frac{3}{4}$ minutes.

While hypnotic dreams have not been demonstrated to be the same as night dreams, several investigators believe there is enough similarity between them to make comparisons. Klein (62) reported that hypnotic dreams lasted about 1 minute. Schroetter (96, 97) has hypnotized subjects to raise their hands at the beginning and at the end of induced hypnotic dreams. In this manner he found their dreams lasted from 1 minute and 20 seconds to 4 minutes and 5 seconds. Using the same method Sweetland and Quay (105) found the length of induced hypnotic dreams averaged about 10 to 20 seconds, the range being from 1 second to 5 minutes. Welch (116) demonstrated that hypnotic dreams could vary according to instructions. The length of a hypnotic dream apparently reflects the nature of instructions given by the hypnotist and the degree of suggestibility of the subject. While these experiments may not give a valid measure of the length of dreaming, they may provide some clue to the speed of imagery.

INCIDENCE AND FREQUENCY OF DREAMING

During night sleep. Several investigators have attempted to determine the frequency of dreams during various parts of the sleeping period. Calkins (13), DeSanctis (22), Bentley (7), and Berrien (8) awakened their subjects at various hours during the night and asked them if they were dreaming. The results of these

studies reveal that more dreams are recalled during the hours just before morning waking and fewest are reported during the middle portion of the night's sleep. There was no period during the night when dreams were never recalled. Berrien reported that sudden awakening of the sleeper resulted in the recall of a greater number of dreams than more gentle techniques, such as whispering to the subject. Calkins (13), Weed and Hallam (115), and DeSanctis (22) noted that the dreams that occurred nearer the morning were much more vivid than those occurring earlier in the night.

Over a period of time. The question as to the percentage of individuals in a group who report dreaming from night to night has also been investigated. Kleitman, Mullin, Cooperman, and Titelbaum (63) found that some of their subjects only rarely dreamed, others dreamed practically every night; but for the total group about 50 per cent dreamed each night. Andress (3) in his study of 49 students for 29 consecutive nights found that each night about 44 per cent reported dreams, the persons dreaming differing from night to night.

Age

Frequently in the literature there is mention of the belief that the frequency of dreaming varies with the age of the subjects. The data on this point are very limited, but they do seem to support a few generalizations.

Children. Since children during the first few years of life cannot make verbal reports regarding dreams, there is no way as yet to investigate the occurrence or frequency of the phenomena. Investigators who have attempted to investigate dreams of children report that their studies are complicated by the fact that the young children often do not separate daytime fantasies from dreams, and that they distort dream reports by the incorporation of daytime experiences. Some workers find that children are reluctant to report on dreams, and others report exactly the opposite. Investigators have obtained dream reports by direct questioning of children, from reports of parents, and by indirect means such as noting dream material that emerges in play.

Claims have been made for evidence of dreaming in children of even less than one year of age. Erickson (28) contended that he observed a pattern of behavior in an eight-month-old infant and followed its continuation for thirteen months, at which age he was then able to elicit a report of an accompanying dream. Several investigators are confident that they have obtained evidence of dream material from children between the ages of two and three.

From reports now available it is indicated that not all children dream, and that there are considerable individual differences in the frequency of dreaming. Blanchard (11) reported that 5 per cent of 230 children under 18 years of age seen at a child guidance clinic could not recall dreaming; another 13 per cent recalled dreaming but were unable to make any report on the nature of their dreams. Jersild, Markey, and Jersild (56) studied dreams of 400 children from 5 to 12 years of age and found, as did Blanchard, that slightly over 5 per cent of them did not recall dreaming. Witty and Kopel (120) interviewed 3,394 boys and girls from kindergarten to the eighth grade, and of these 16 per cent reported never dreaming, 63 per cent sometimes, and

21 per cent often. Blanchard (11) found a statistically reliable difference for more dreams recalled after age six than before, but she believed this difference resulted from the better ability of older children to recall and to report dreams. Thus the evidence at hand so far indicates that probably 5 to 15 per cent of children cannot recall dreaming.

Adults. Manacéine (73) and Heerwagen (48) contended from their investigations that young adults dream more frequently than the aged. Heerwagen found that the greatest number of dreams occur between the ages of 20 and 25. Hamilton (45) reported for 100 males that erotic dreams increased rapidly in number after age 11 for males until about 80 per cent reported such experiences by age 20. For the 100 females in his study age did not show such a relationship. Kinsey, Pomeroy, and Martin (61) also found that approximately 83 per cent of 5,300 adult males included in their report had experienced sex dreams which terminated in nocturnal orgasm. The frequency of such dreams decreased steadily in number after age 30.

Intelligence

The relationship of intelligence to frequency of dreaming has received only slight attention. In a study of children Despert (23) found the greatest number of dreams was reported by those with the highest intelligence quotient (IQ), the smallest by those with the lowest IQ. Not all children with high IQ's, however, reported a high frequency of dreams. Jersild *et al.* (56) reported contradictory findings in that the children in their study with IQ's over 120 dreamed less frequently than those with IQ's below 120. Blanchard (11), Kimmins (60), and Selling (99) found no relationship between intelligence and frequency of dreaming.

Very few data bear upon intelligence and frequency of dreaming among adult populations. Manacéine (73) collected dream reports for five years from 37 persons of varying intelligence. He concluded that the number of dreams is less in persons with low intellectual capacity than for those with greater endowment.

The whole question of intelligence and frequency of dreaming has not been adequately studied. Future investigators of this topic should give more attention to: (*a*) the problem of the validity and reliability of the intelligence measure used, (*b*) selection and description of the population studied, and (*c*) obtaining more detailed information regarding the nature of dreams such as richness and variety of content. It appears to the reviewer that the dream life of individuals should reflect the range, richness, and variety of mental activities which are associated with the higher levels of intellectual capacity.

Sex Differences

The question of sex differences in dreams has long been debated, but little systematic evidence has been collected on the topic. A survey of the studies on dreaming revealed that data were often reported without any mention of the sex of the subjects. In those few studies in which the results were analyzed according to sex, it was found that women reported a higher frequency of dreaming than men. This sex difference was cited by Chrysanthis (15) in a report on 1,408 students from 12 to 18 years of age, and by Middleton (77) for 277 college men and women. However, in another study on a similar population Middleton (76) found contradictory results. Jastrow

(54) found that blind women dream more than blind men. Heerwagen (48) noted in his data that unmarried women reported a greater number of dreams than did married women.

The evidence for sex differences in frequency of dreaming among children is contradictory. Foster and Anderson (31) found no reliable sex differences in their investigation of the number of unpleasant dreams of 519 children. On the other hand, Witty and Kopel (120), in their report of 3,394 children, found that girls reported a slightly greater number of dreams than did boys. Jersild et al. (56) found no sex differences in the number of pleasant or unpleasant dreams.

Sex differences in the content of dreams have also been the subject of a few investigations. Cason (14) and Schubert and Wagner (98) reported that girls dream more frequently of home, family, and friends than boys do. Boys had far more fear dreams than girls, according to Kimmins (60). Gahagan (37) reported that the only significant differences found between the dreams of 228 university men and 331 university women were: (a) a greater percentage of women had "physical examination dreams," and (b) men had a greater percentage of dreams of being nude. Middleton (76) also found men dream of being nude more often than women do.

Husband (52) studied the dream reports obtained from 25 university males and 25 university females which he obtained by personal interview. His results indicated that women's dreams were much more vivid and more emotional than men's. Women more frequently transferred their worries, love, and problems into their dreams than did the men. Women also dreamed more about their "boy friends," while men

did not dream so often about their "girl friends." Hall (44), in an unsystematic report on over 10,000 dreams, found that women dreamed twice as often about males as females, whereas men dreamed equally about males and females. Middleton (76) found that women discussed their dreams with others twice as often as did the men. Both Husband (52) and Middleton (76) found sex differences in frequency and content of erotic dreams.

CONTENT OF DREAMS

The content of dreams has been of primary concern to many investigators but most of their reports are descriptive in nature. Among the earlier studies on content of dreams many were concerned with specific types of content such as dreams of falling, levitation, premonition, dreams of the dead, dreams within dreams, etc. Investigators simply looked for and found examples of dream content that illustrated a particular type. In their reports on such dreams they seldom provided any systematic data as to incidence or frequency of such dreams in a given population, or presented any relationship evidence. In general most of such work is illustrative of the content that can be found in dreams, but it provides little data on which systematic studies can be based.

Practically all investigators emphasize in one way or another the congruity and continuity between daily experiences and fantasies of the subjects, and the content of the reported dreams. For example Weed and Hallam stated, "The existence of some connection between the dream world and the waking world, that is, the suggestion of the dream world by some actual experience, is traced in most of the dreams" (44, p. 409). Calkins

(13) traced 90 per cent of the content of 375 dreams of two adults to events in their waking life. She found that many more of the dreams were connected to the more unimportant recent events in their lives than to the important ones. Andrews (4) similarly found that nearly 90 per cent of 118 of her dreams were clearly suggested by waking experiences. Middleton (76) reported that in 51 per cent of the dreams of 170 college students there were some elements traceable to the previous day's experiences.

Among the investigators who have studied the content of children's dreams, a similar generalization is reported. Jersild *et al.* stated, "The content of dreams is likely to reflect any experience, wish, fear, fancy, or circumstance which occurs in the child's waking moments, and dreams are likely to reflect unpleasant events in the child's experience somewhat more frequently than pleasant ones" (56, p. 133). Witty and Kopel made a similar statement based on their investigation (120, p. 205).

The content of children's dreams has been classified by several investigators and general agreement is noted in the findings even though different systems of classification have been used. Blanchard (11) classified the dreams of 230 children and found in order of frequency the following content: parents, animals, fear incidences, play activities, robbers, and death. She found no relationship between content categories and the individual's IQ, mental age, or chronological age. Witty and Kopel (120) reported similar findings. Jersild *et al.* (56) classified the content of "bad" and "good" dreams. Among the former were such content categories as vicious animals, "bad" people, physical distress and injury,

supernatural creatures, fires, storms, falling, being chased, etc. "Good" dreams included acquiring such things as toys, food, money, clothes, pets, and other content such as travel, play, visits with companions, etc. Despert (23), in a study of the dreams reported by 39 children who were clinical patients, classified the content of their dreams as follows: 75 human beings, 55 animals, and 60 inanimate objects. She found the better adjusted children in this group had more human content in their dreams than those who were less well adjusted. Kimmins (60) makes several comments on dream content and its relationship to age, health, etc. but gives no actual data. For example she states that children from poorer districts were found to dream more frequently about toys, food, clothes, etc. than did those from well-to-do districts.

The content of nightmare dreams of children was studied by Cason (14). He reported the most frequent contents were: animals, death, murder, being chased, falling, accidents, and other unpleasant experiences. He found no relationship between content and age, sex, and education. Wile (119) traced the content of dreams of 25 children who during their sleep yelled, cried, walked, etc., and found that the dreams connected with these disturbances were based upon actual experiences and not born of fantasy. Usually they were related to recent events which involved physical injury or punishment, or were related to psychological trauma.

Dreams of special groups. There has been a long-standing belief that special groups, such as criminals, psychopaths, prostitutes, the mentally deficient, psychotics, etc., have characteristic patterns of content in their dreams. Most of the reports on

these various groups, even when based on actual case studies, are so inadequately reported and unsystematic that little confidence can be placed in them. For example, DeSanctis (20, 21) has made voluminous reports on the dreams of several special groups, but his work is certainly not objective or systematic. His reported findings need to be checked by other investigators.

The content of the dreams of criminals has received some brief attention. Selling (99) obtained by interview a report on the dreams of 200 juvenile delinquents and 100 convicts. He found that 80 per cent of the dreams referred to home and activities related to it. Hanks (46) secured a report on one dream from each of 50 convicts and found that returning home was the most frequent content topic.

Stability of content. Hall (43) attempted to study the stability of dream content. He classified the dreams reported in the diary of an adult male for a two-year period and then compared the dream categories for the first year against the second year. He found a high degree of consistency in the content categories and interpreted the finding as reflecting the stability of this adult's personality dynamics.

Erotic dreams. Erotic or sex dreams have been given frequent and lengthy descriptive consideration in the past, but very little quantitative data bear upon the topic. In considering reports made on sex dreams, attention must be given to the possible effects of the socially disapproved nature of the topic upon the validity of the individual's report. For example, Foster and Anderson (31) and Despert (23) stated that the children in their respective studies did not report any erotic or sex dreams. This is not surprising since they did not ask about such dreams and depended solely upon the spontaneous reports of the children upon a socially disapproved topic. The fact that certain children do experience erotic dreams was indicated by Blanchard (11), who found that 9 per cent of 230 children in her group reported their occurrence. A systematic study of erotic dreams of children, however, has not yet been made.

In considering erotic dreams of adults an additional problem is encountered because reporters usually fail to distinguish between data bearing upon erotic dreams and nocturnal orgasm. Dreams which accompany sexual orgasm are usually erotic in content, but may be composed entirely of other content. Also, orgasm may occur without the recall of any accompanying dream. Therefore the data for each phenomenon must not be considered as identical, as has been the assumption in some of the reports.

Hamilton (45) explored erotic dreams and sexual orgasm in a group of 100 males and 100 females. His data, while systematically gathered, were presented in such a manner that comparisons and interpretations based on them are often very difficult to make. But his evidence indicated that before age 20 about 75 per cent of the men had erotic dreams which accompanied orgasm, while only 3 per cent of the women reported such experiences before age 20. About 90 per cent of the men reported dreams associated with orgasm during a lifetime, while only 33 per cent of the females made a similar report. In most cases women who did experience erotic dreams and orgasm did so only after first sexual intercourse.

Middleton (76) stated that college men in his study reported twice as many erotic dreams as did the

women. Selling (99) found that among 200 juvenile delinquents 33 per cent reported erotic dreams, and among 100 adult convicts 85 per cent made a similar report. Usually erotic dreams were accompanied by nocturnal emissions in the adult group. Franklin, Schiele, Brožek, and Keys (33) and also Miles (78) noted a reduction or disappearance of sex dreams among subjects undergoing experimental semistarvation and rehabilitation.

Husband (52), in a very carefully conducted study, obtained data by means of personal interview on the dreams of 25 college males and 25 college females. He included several questions about erotic dreams. He found that all the males had experienced at least one erotic dream within the last year, and all but two of the females made a similar report. About 50 per cent of the subjects reported sex dreams which involved some major type of sexual activity. The sex dreams of the females were more frequently about a relationship with a close friend, while those of the males were more often about casual companions. The females reported more sex dreams when their love affairs were at a height, but for males there was no such relationship. Females had more sex dreams when their closest friend or spouse was absent, but this was not true for the males. Dreaming of "unnatural" sex relationships was reported by 14 of the 25 females and 18 of the 25 males. Sexual contacts in dreams were made with animals, children, homosexual partners, mother, sister, etc. This study shows how much socially tabooed information can be obtained by use of the personal interview when confidence in the report and the research has been established by the investigator.

EXTERNAL STIMULATION AND DREAMS

From ancient times individuals have reported the influence of accidentally or intentionally introduced stimuli upon the course of dreams of the sleeper. Most individuals can recall how some light, sound, odor, or presssure, etc. has influenced the nature of his dreams. The literature of the past centuries is replete with many interesting cases of the effects of external stimulation upon the course of dreaming. Frequently cited examples include dreams of freezing when blankets fall off during the night, dreams of fire when lightning falls upon the eyelids, dreams with olfactory components when the sleeper is in a smoke-filled room, etc.

In the last century several attempts were made to determine the effect of experimentally introduced stimuli upon the course of dreams. Historically this was the first type of experimental work done on dreams. Apparently the first experiments of this nature were conducted by Maury (74) in 1861. During sleep he had an assistant tickle his face, pinch his neck, open perfume bottles near his nose, drop water on his forehead, etc. When awakened, he recorded the nature of his dreams. He found in many cases that his dreams showed the influence of the introduced stimulus. Several investigators since then have performed similar experiments and have obtained like results. Early experimenters who have been cited as studying the effect of external stimuli upon the course of dreams include Hildebrant (49), Ladd (64), Weed and Hallam (115), Clavière (16), Vaschide (110), DeSanctis (22), Vold (111, 112), Straeke (104), Stepanow (103), Poetzl (88), and Cubberly (17). A few examples from these studies will suffice to show their gen-

eral nature. Vold (112) conducted a series of experiments involving immobilization of extremities by means of bandages, gloves, etc., and also a series in which he moved the limbs and body position of the sleeper. He found that such techniques did have an influence upon the reported dream pattern. Stepanow (103) transmitted a variety of musical stimuli through a speaking tube which was connected with the sleeper's room. He made a report on the influence of the music on the nature of dreams. Cubberly (17) pasted small gummed labels, called "tensors," on various parts of the body after the subject had gone to sleep. He found that in approximately 95 per cent of the experiments the "tensor" stimulus had an effect upon the course of the dreams. Usually the stimulus became the core of the dream content. All these studies showed that external stimuli, if sufficiently intense, did have a definite influence upon the course of dreams during normal sleep.

A more controlled experiment was performed by Max (75), who studied action currents in the peripheral musculature during sleep of deaf-mutes. He reported that the onset of dreams could be detected in most instances by the appearance of large action-current responses in the arm and finger muscles. When awakened when such responses were evident, in practically all instances the subjects stated that they had been dreaming. The same subjects, when awakened at other times, usually reported they were not dreaming. He also induced dreams experimentally by external stimuli and noted accompanying action currents. The maximum action-current response connected with a dream lasted for $2\frac{3}{4}$ minutes. Normal subjects studied in the same way did not produce action currents

during dreaming similar to those found for the deaf-mutes.

A study of the influence of external stimuli upon hypnotic dreams was carried out by Klein (62). He used ten different stimuli such as perfume, asafetida, sounds from a tuning fork, etc. to induce hypnotic dreams in a series of experiments with eight adult subjects. He found that hypnotic dreams could be induced by the various stimuli used, and that the content of the dream reflected the type of stimulus employed. The subjects were able to recall faithfully the hypnotic dreams, and they considered them as indistinguishable from night dreams. "Falling dreams" were induced by lowering or raising the body, or parts of it, when the hypnotized subject was at rest in a bed.

PERSONALITY AND DREAMS

Since ancient times there has existed the belief that dreams provide a basis for understanding and interpreting personality. Freud (34) presented the first major treatment of this topic and supported his theorizing with case materials. Today many psychiatric and psychological practitioners claim that dreams provide useful data for reaching a better understanding of personality dynamics and problems of patients. It should be pointed out that most of their observations and opinions about dreams and personality relationships have not been subjected to systematic and quantitative investigation. However, many hypotheses have emerged from their observations; some of them have been subjected to verifiable procedures and in some cases have yielded valuable data.

In survey, the studies on dreams and personality were usually found to be rather segmental investigations

and not too closely related to one another. Therefore comparison between findings from different studies cannot be easily made.

Emotions and Dreams

The emotional tone of dreams has been a subject of much discussion and a few investigations. Apparently dreams do have a very close relationship to the emotional life of the dreamer. Weed and Hallam (115) found that 92 per cent of the 381 dreams of six subjects were considered by these subjects as having a definite emotional tone. Other investigators have secured reports which testify to the emotional nature of most dreams, and in addition they have attempted to analyze this element of dream phenomena. For example, several investigators have inquired regarding the pleasant-unpleasant tone of dreams.

Pleasant-unpleasant tone. Apparently individuals seldom have difficulty in classifying the affective tone of dreams as either pleasant or unpleasant. Practically all investigators of this aspect of dreaming found that unpleasant dreams outnumbered pleasant ones. The following studies of this topic are representative: Calkins (13) found 75 per cent of 375 dreams of two adults were classified as unpleasant; Weed and Hallam (115) stated that 57 per cent of 381 dreams of six subjects were considered as unpleasant; and Bentley (7) reported that 70 per cent of 54 dreams of five adults were unpleasant. Lovett (68) found that unpleasant dreams were more frequent than pleasant ones in both a psychiatric group and in a normal control group. Large individual differences have been noted in the reports on the frequency of pleasant-unpleasant dreams. Some individuals report

that practically all their dreams are pleasant; others find most of theirs unpleasant. Apparently no sex differences occur in the frequency of pleasant-unpleasant dreams according to the data collected by Gahagan (37) on 559 university students.

Children's dreams were also found to be more frequently unpleasant than pleasant in the reports made by Jersild *et al.* (56), Foster and Anderson (31), Witty and Kopel (120), and Despert (23). The content of unpleasant dreams appeared to correspond more closely to the children's waking fears than to their actual daily experiences. Unpleasant dreams were found to have a greater tendency to recur than did other dreams. Jersild (56) and Witty and Kopel (120) did not find much change in the frequency of unpleasant dreams with increasing age. On the other hand, Foster and Anderson (31) found their frequency to decrease with age. These authors also reported that unpleasant dreams appear to increase during periods of illness, overexcitement, fatigue, emotional upsets, and physiological disturbances. Despert (23) failed to discover any relationship between health and the frequency of unpleasant dreams. Foster and Anderson (31) noted that children who slept alone had fewer unpleasant dreams, and in addition that the number of siblings in the family was not related to the frequency of unpleasant dreams. Jersild *et al.* (56) found that children with IQ's of over 120 reported more unpleasant dreams than those with lower IQ's.

Emotional perseveration. The continuation of the emotional tone of dreams into waking hours has been considered by a few investigators. Jersild *et al.* (56) stated that almost one-half of the children in their study were so disturbed by unpleasant

dreams that they expressed a wish that they would never dream again. Pintner and Lev (87) found that two-thirds of the 270 boys and 270 girls in their study had daytime worries about unpleasant night dreams. Middleton (76) reported that 45 per cent of the 270 adults stated that unpleasant dreams produced prolonged moods in subsequent waking periods.

Bagby (5) studied the effect of emotional experiences during waking hours upon dreaming. He examined a few dreams reported by an engaged couple while under "sexual stress," and concluded that the emotional stress in the waking life of these two was reflected in their dreams.

Emotional stability. An old and rather popular belief is that individuals who are emotionally unstable dream more frequently than those who have a more stable emotional life. Berrien (8) attempted to study this hypothesis in his 1930 investigation. He administered the Colgate Mental Hygiene Test to four subjects to obtain emotional stability-unstability ratings. He found in this small group that the emotionally unstable reported more dreams. In 1933 he conducted a similar study (9) in which he administered the Thurstone Personality Test and the Colgate B2 Psychoneurotic Scale to 81 university students and thereby derived a stability-unstability rating for each one. An analysis of these ratings and reported dream frequencies gave little or no support to the belief that the emotionally unstable had a greater number of dreams.

The frequency of dreaming among children classified as anxious and non-anxious was reported by Despert (23). From her study of 190 dreams obtained from 39 children ranging in age from two to five who were clinic patients, she concluded that all children who had frequent dreams were among the more anxious children. Not all anxious children, however, reported a high frequency of dreams.

Psychiatric Groups and Dreams

Countless contentions have been made that the dream life of various psychiatric groups has characteristic patterns. Among the many current notions regarding the dream life of particular groups can be found such beliefs as: psychopaths never dream, hysterics have more erotic dreams, prostitutes and criminals seldom dream, psychotics experience more color in dreams, etc. It is quite possible that certain psychiatric groups, because of the nature of their disorder, might have certain unique characteristics to their dream life. At present, however, there is very little verifiable data supporting the contentions. Fragmentary evidence has been published regarding the dreams of the feebleminded (114), delinquent (20), epileptic (30), and other populations. The findings, however, are not very convincing.

Lovett (68) studied the dreams of 293 psychiatric patients and 260 normals for similarities and differences. He found no significant differences in the sexual content of dreams as reported by the two groups. The members of the psychiatric group were found to have more human content in their dreams, more incidents of fear and anxiety, and greater difficulty in separating fantasy from dream material than did members of the normal group. A study of dreams of 12 schizophrenics was made by Kant (57). He examined 200 dreams reported by the patients and concluded that schizophrenic patients did not dream less than normals, and that no dreams appeared which were

peculiar to the schizophrenic disorder. The psychiatric group did attribute more significance to their dreams than did the normal group. The more recent reports of Trapp (109) and Noble (85) in this area are consistent with those previously cited.

Alexander and Wilson (1) made an attempt to relate categories of dream content with the medical diagnosis of psychosomatic patients. The dreams of 18 patients were classified into such categories as "inhibited receptive," "satisfied receptive," etc. A study was then made of the frequency of each dream category reported by patients with peptic ulcer, chronic diarrhea, constipation, etc. In general, findings were in agreement with current psychoanalytic theories of psychosomatic disorders.

Traumatic Experiences and Dreams

Many cases are cited in the literature to show the effect of traumatic experiences upon dreaming. Case reports given by Anderson (2), Bagby (5), and Gahagan (37) are illustrative. The traumatic experiences of war and their relationship to dream life are discussed by Grinker and Spiegel (40), Kardiner (58), and several others.

Group studies of traumatic experiences and possible relationships to dreaming are very few. Wile (119) selected a group of children who were experiencing violent dreams and then searched for traumatic experiences in daily life which might explain the disturbing dream life. In his study he selected 13 girls and 12 boys who cried, yelled, and fought during their sleep. He found that the violent dreams of these children appeared to be related to their actual experiences. In about 20 per cent of the cases the dreams were related to physical injury or punishment, while

80 per cent were centered around traumatic psychological experiences.

Visual Perception and Dreams

A few attempts have been made to study the effect of the presentation of types of visual stimuli during waking periods upon subsequent dream material. Malamud and Linder (72) cite such an experiment conducted by Poetzl (88) in 1918. He had pictures presented tachistoscopically for 0.1 second to normal subjects and then asked them to report what they saw. Then they were told to make a record of any dream they had about the pictures. The subjects brought out more details of the pictures in the dream reports than were mentioned after viewing the pictures. He explained the difference in immediate recall and in dreams as a function of personality dynamics.

Vold (111) gazed at small plastic objects or figures before going to sleep and then studied his dreams in order to detect what influence they had upon their course and content He reported that very seldom did the form, size, or color of the object remain unchanged, but that the object often appeared in some alternative or distorted manner.

Malamud and Linder (72) compared the recall of visual materials immediately after presentation with the dream reports of the same stimulus. They presented a series of pictures, one at a time, for 30 seconds to 27 male and 20 female psychiatric patients. After each presentation of a picture a five-minute period of conversation intervened. The subject was then asked for a complete description of the picture. The patient was then told to dream of the picture and, if he did, to write a report of it. The investigators found that some content which did not appear in the

immediate recall after presentation did appear in subsequent dream reports. They contended that the details recalled or omitted were related to the individual's early personality development and the current psychiatric disorder.

Sarason (92) compared the Thematic Apperception Test stories obtained from 25 mentally defective females with dreams reported by each individual. He reported that a similarity was found between the two sets of data. He pointed out that not in every case were all the major themas in the thematic materials found also in the dreams, but that in no case were data from the two sources at complete variance. Griffiths (39) found a striking similarity between usual dream images of subjects and responses obtained from inkblot reactions.

Ego Involvement and Dreams

Sweetland and Quay (105) investigated the influence of "conceptual" stimuli upon the nature of hypnotic dreams. They prepared a list of 40 ego-involving and 10 neutral statements. These were presented, one at a time, to hypnotized subjects with the instructions "to dream" to each one. For 16 college subjects they found that reaction time and duration of dreams had no relationship to the ego-involving stimuli. The data did show, however, that each subject had his own characteristic speed and duration of dreams. Recall of dream material was found to be a function of recency, and to be independent of the nature of the stimulus.

Symbolization. Freudian theory (34) gives considerable attention to the symbolization which occurs in dreams and the interpretation of such materials. A few investigations of symbolization in hypnotic dreams have been conducted, and the results may throw some light on symbolization which occurs in night dreams. Schroetter (96), as early as 1911, studied the effect of direct and indirect representation of suggestions upon the nature of hypnotic dreams. He prepared a series of statements which suggested wish fulfillment, clang associations, abnormally small dimensions of dream figures, and various types of sexual activity. These statements were read one at a time to each hypnotized subject. After each presentation the subject was instructed "to dream" about it. A report is given of the dream material obtained from such experiments. The investigator points out the symbolic representations in the dream productions. Negative results were probably encountered but they are not reported. Roffenstein (90) conducted a very similar investigation. He instructed a 28-year-old, uneducated female to dream symbolically while under hypnosis of fellatio, homosexual intercourse, rape, and intercourse with her father. The dream material obtained is reported. The investigator points out symbolization in the dream content and discusses its symbolic significance.

Symbolization or alteration of stimulus material in hypnotic dreams and its relationship to emotional adjustment of the dreamer were reported by Sweetland and Quay (105). They presented 16 college students with a series of ego-involving statements as topics for hypnotic dreams. The resulting dreams were classified into symbolic and nonsymbolic categories. The best adjusted subjects, as indicated by two psychological tests, had the greatest number of symbolic dreams. Only a slight relationship appeared between intelligence and

frequency of symbolization.

Nachmansohn (84) attempted to study dream symbolization, dream censorship, and dreams as a protector of sleep by use of the posthypnotic suggestion "to dream" during regular night sleep. He prepared a series of statements which served as stimuli for induced hypnotic dreams. After reading a statement to the hypnotized subject he instructed him "to dream" in a symbolic manner to the stimulus. Censorship in dreams was similarly investigated. Also he attempted to show how dreams act as a protector of sleep by a demonstration using the posthypnotic suggestion "to dream" during night sleep. He reports on a subject who had been awakening each night from a recurrent headache. He gave the subject a posthypnotic suggestion "to dream" at night of the headache. He claimed that as long as the posthypnotic suggestion was used, the patient did not awaken at night.

Farber and Fisher (29) studied symbolization occurring in hypnotically induced dreams by using both sexual and nonsexual suggestions for dream stimuli. They pointed out the wide range that occurred in the nature of the symbolization. They argued against a too narrow interpretation of the process.

PHYSIOLOGY AND DREAMS

Physiological studies bearing on dreaming can be separated for discussion purposes into two groups. In the first group are placed those investigations which report data bearing upon some physiological activity that occurs while dreaming is in process. Considered under this grouping are such phenomena as brain potentials, action currents, galvanic skin responses, blood pressure, etc. The second group includes studies which report that certain physiological conditions have an influence upon the nature and content of dreams.

Physiological Activity during Dreaming

One major problem encountered in physiological studies is establishing the fact that the sleeper is or is not dreaming. Investigators have had to rely upon awakening subjects at times when selected physiological data appeared and asking the subject if he was dreaming. Using this method a few studies have been made which provide some data bearing upon physiological activity during dreaming.

Brain potentials. Loomis, Harvey, and Hobart (65, 66, 67) and Davis, Davis, Loomis, Harvey and Hobart (19) obtained records of brain potentials during sleep. These were analyzed according to reports of the subject as to whether or not dreaming had occurred prior to awakening. Loomis *et al.* reported: "We are now inclined to believe dreams are not likely to occur with any unusual pattern of electrical potentials but with a state of sleep" (66, p. 142). Davis *et al.* make a similar report: "We have no evidence relating dreams to specific changes in the record" (19, p. 33). They reported that dreaming does occur in both the B state of electroencephalogram (EEG) recordings (low voltage, and alpha rhythm is lost) and C state (spindles: short groups of 14-per-second waves appear and also random "delta" waves 0.2 second or more in length). Concerning the C and D states they have no report.

Blake and Gerard (10) found the presence of delta waves to be related either to dreamless sleep or to the inability to recall having dreamed.

For one subject it was found in nine-tenths of the trials that if alpha waves were missing for nine seconds and no delta waves were present, the subject could remember a dream when awakened. Sirna (100) made EEG recordings of hypnotically induced dreams and found that the physiological changes during hypnotic dreams show no positive correlations with physiological changes to be seen in the normal sleeping dream.

Muscular action currents. As indicated previously, Max (75) found an interesting relationship between action-current records taken from the peripheral musculature of sleeping deaf subjects and the reports, gathered on awaking, of the presence or absence of dreams. The onset of dreams could be detected in most instances by the appearance of large action-current responses in the arm and finger muscles. Such currents were usually unaccompanied by overt muscular movements. Also, the author investigated hypnotic dreams which were induced by sensory stimulation. The same type of accompanying action currents were noted as appeared during night dreams. The investigation was repeated using normal subjects, but negative results were obtained.

Galvanic skin responses. No studies of galvanic skin responses during night dreaming were found. Ikin, Pear, and Thouless (53) investigated galvanic responses made by waking subjects when asked to free-associate to materials drawn from their own dreams. The data showed that memories of emotional dream materials did cause noticeable GSR's. In one case recall of a dream which occurred over 15 years before produced a definite response.

Circulatory activity. MacWilliam (71) studied blood pressure and heart

action in sleep and dreaming. He found that the systolic blood pressure could rise during a dream from 130 to over 200 mm. He further stated that while the most striking cardiovascular effects are found in dreams with strong emotional content, vivid dreams of active movement (such as cycling) may be associated with a pronounced rise in blood pressure.

Respiratory activity. Rowland (91) obtained pneumographic records while subjects under hypnosis were instructed to dream of a series of emotional rated words. Each word had previously been rated by judges as to the degree of emotional value it usually carried. The pneumographic records showed that the increase in irregularity of breathing was roughly proportional to the scaled value of each word.

Gastric contractions. The relationship between hunger contractions of the stomach and reports on dreaming have shown a positive correlation in the reports made by Wada (113) and Scantlebury, Frick, and Patterson (93, 94). The stomach movements in these studies were studied by use of a gastric balloon. At certain times during hunger contractions and quiescence, subjects were awakened and asked if they were dreaming. The evidence showed that dreams occur only in the period of active hunger contractions. No dreams were found during the quiescent period. Scantlebury et al. (94) also investigated the effects of hypnotically induced dreams and found that when food was a part of the dream content it had an inhibitory influence upon stomach movements. McGlade (70) provided evidence to show that dreaming occurred at the end of a series of twitches in peripheral members of the body which accompanied the evacuation of the stomach.

Physiological Conditions Influencing Dreaming

Several writers contend that certain physiological tensions, disequilibria, or dysfunctioning influence the nature and content of dreams. Claims are made that such conditions as hunger, thirst, fatigue, sexual deprivation, bladder and bowel distention, respiratory and circulatory irregularities, organic diseases, etc. produce alterations or special characteristics in dream patterns. Other factors falling into this topic are the effects of drugs, hormones, narcotics, etc., upon dreaming. The amount of evidence supporting many of these beliefs is meager or totally lacking.

Hunger and dreams. The belief that hunger or periods of starvation increase the frequency and content of dreams pertaining to food and eating is frequently met. There is some support for the belief from a few studies. Several anthropologists have reported that members of tribes who are exposed frequently to hunger have at such times an unusually large number of dreams about food gathering, preparation, and eating. An example of such a report is made by Holmberg (51) in his study of the Siriono Indians.

Benedict, Miles, Roth, and Smith (6) and Sorokin (102) investigated the dreams of hungry persons and reported that they are primarily centered around food and eating. However, in both of these studies the methods of investigating frequency of dreaming were so questionable that little confidence can be placed in their reports.

Contradictory reports on the influence of hunger and starvation upon dream patterns are found in the University of Minnesota starvation experiment carried out by Keys, Brožek, Henschel, and Mickelson (59). In this more carefully conducted study a report was obtained periodically from 36 men who were undergoing a starvation experiment. This study extended over a period of 20 weeks. The investigators found that dreams of food were rare, and no marked tendency was noted for them to increase during the period of starvation or during the period of rehabilitation which followed.

As reported earlier, Miles (78) and Franklin *et al.* (33) reported that sex dreams were greatly attenuated or entirely eliminated when subjects were undergoing reduction in normal diet or under semistarvation experiments.

Anoxia. Lovett (68) cites a study by Monge (79), who reported an increase of disturbing dreams when dwellers of high altitudes suffer from chronic mountain sickness. McFarland (69) reported on dreaming among members of an expedition going from sea level to high altitudes in the Andes. The generalizations made from their reports were that while ascending, dreams increased in frequency, were unusually fantastic and illusory, and a shift in content was noted from sex and home to worries about health. But at elevations of over 17,000 feet dreams practically never occurred, even with nocturnal emissions. The conclusions drawn were that the general physiological state appeared to be equally as important as inner conflicts or motives in determining the nature of dreams. Lovett (68) found no significant relationships when he correlated the arterial oxygen saturation expressed in "reduction-time" seconds and the frequency of nightmares, color in dreams, and recall of dreams.

Hormones. Finley (30) reported a case in which endocrine therapy had a marked influence on dreaming. When the subject was being given one grain of the extract of whole pituitary gland daily, there was a large increase in the frequency of her dreams. In addition the dreams were much more vivid, pleasant, full of action, and included many colored ones. Increase in ability to recall dreams under these conditions was also noted. When the patient was changed to suprarenal gland therapy fewer dreams were reported, and most of them were unpleasant. There was no color in the dreams and recall was difficult. Under this hormone she would awaken at night from startling dreams and would exhibit considerable emotional disturbance. At the termination of therapy she returned to her normal dream pattern. The patient reported an increase in the number of pleasant dreams around her menstrual period. Daniels (19) studied the occurrence of erotic dreams of three females during the menstrual month and found that, while such dreams appeared throughout the month, they showed a concentration around the supposed time of ovulation.

Sensory disturbance. Hoff and Pötzl (50) reported on the dreams of two subjects who had central and labyrinthine disturbances. An unusually large number of their dreams were of falling and flying. Grünbaum (41) and Eisinger and Schilder (26) cited cases in which a tumor or organic involvement of the auditory nerve produced a high incidence of auditory dreams.

Enuresis. Investigators (83) who have studied the dreams of enuretic children have reported that these children often experience vivid and realistic "toilet dreams" just before or during the act of urination. These reporters have failed, however, to show any difference in such dreams of enuretics and normal children. Many children and adults also experience "toilet dreams," but do awaken before urination.

Narcotics and drugs. Many writers have discussed the production of hallucinatory imagery by use of drugs, anesthetics, narcotics, and other pharmaceutical agents. Most of this material, however, makes no mention of their influence upon night dreams. One study by Jelliffe (55) reported the increase of color in night dreams when the subjects were under the influence of morphine. Another author mentioned that drug addicts, when undergoing withdrawal treatment, dream frequently of seeking narcotics. The question of the influence of drugs and narcotics upon night dreaming, however, has not been systematically investigated.

EVALUATION OF RESEARCH

There are only a few research articles on dreaming which have been so designed and reported that they can be repeated and published findings thereby checked. Future studies on this topic need to give greater consideration to the simple research criterion of repeatability if systematic knowledge of dreaming is to be extended. In the past most of the studies on dreaming have been limited to case reports and speculations. While these have some value in themselves, they should lead eventually to more quantitative types of investigations. This goal, however, has not been served by most of the past studies.

A very noticeable weakness in many of the past investigations is the quality of scientific reporting. Often only findings or generalizations are

reported even though they are based on a respectable body of data. Unless essential basic data are presented, future research workers will be unable to verify and extend past results. Presenting basic data also makes it possible for future researchers to separate actual findings from interpretations. Too frequently a reviewer of the studies on dreaming simply cannot tell whether statements made by an investigator are based on fact or are a product of his speculations.

Investigations of dreaming are particularly weak when it comes to describing the nature of the population on which they are based. Many investigators mention only the number of subjects involved in a study and seldom add any information about sex, age, level of intelligence, health, economic status, education, etc. Yet the literature is full of speculations as to the influence of such variables upon the nature and content of dreams. Many of the studies have been based upon a very limited or select group of subjects. Findings reported from such studies should be considered as tentative or exploratory. Too often they are treated as definitive or used as the basis for generalizations which are not justified.

Another weakness is noted in the failure to employ control groups when comparative statements are made about one or more selected groups. In the literature claims are frequently made that particular groups such as the feebleminded, prostitutes, criminals, psychotics, the aged, etc. have characteristic dreams. Seldom do such studies provide any comparative data from a control group. Therefore claimed characteristics of groups are not adequately established.

More precise definitions are needed in selecting and classifying various aspects or characteristics of dreams. Some of the contradictory findings appear to be artifacts of the classificatory treatment of data rather than actual characteristics of dreams. For example, contradictory evidence was found regarding "recurrent dreams." Yet in the studies no definition of the category was given. Other concepts needing definitions so that results can be checked would include such terms as symbolization, censorship, projection, categories for classifying dreams, etc.

A survey of past studies on dreaming probably gives one of the best examples in the field of psychology for the need of statistical controls and treatment of data. Many of the past workers failed to use the statistical procedures and checks available to them. Present day investigators of dreaming have many more and better statistical tools to apply to research on this topic. Statistical tests of validity, reliability, significance of differences, population sampling, etc. must be employed when applicable or the research involved is subject to sharp criticism.

A large portion of the data on dreaming is based on individual interviewing. Such a procedure has inherent and potential weaknesses, but like any method there are degrees of control which can be exercised over many of these factors. The researcher should be aware of possible distortion or bias that might appear in data collected by personal interview. At the same time it should be recognized that the interview method has contributed much valuable data and insight into the phenomenon of dreaming.

The subject of dreaming is considered relevant to the dynamics of

behavior by many practicing psychiatrists and psychologists. Even academic psychologists who are interested in thought and imagery processes find that dreaming is a phenomenon of importance to them. The psychologist with his research orientation and research abilities could undoubtedly help advance the knowledge of dreaming from the present speculative and descriptive stage and give it a more experimental and quantitative basis.

REFERENCES

1. ALEXANDER, F., & WILSON, G. W. Quantitative dream studies; a methodological attempt at quantitative evaluation of psychoanalytical material. *Psychoanal. Quart.*, 1935, 4, 371–407.

2. ANDERSON, J. E. The dream as a re-conditioning process. *J. abnorm. soc. Psychol.*, 1927, 22, 22–25.

3. ANDRESS, J. An investigation of the sleep of normal school children. *J. educ. Psychol.*, 1911, 2, 153–156.

4. ANDREWS, GRACE A. Studies of the dream consciousness. *Amer. J. Psychol.*, 1900, 12, 131–134.

5. BAGBY, E. Dreams during periods of emotional stress. *J. abnorm. soc. Psychol.*, 1930, 25, 289–292.

6. BENEDICT, F. G., MILES, W. R., ROTH, P., & SMITH, H. M. Human vitality and efficiency under prolonged restriction diet. *Carnegie Inst.*, Washington, Publ. No. 280, 1919.

7. BENTLEY, M. The study of dreams. *Amer. J. Psychol.*, 1915, 26, 196–210.

8. BERRIEN, F. K. Recall of dreams during the sleep period. *J. abnorm. soc. Psychol.*, 1930, 25, 110–114.

9. BERRIEN, F. K. A statistical study of dreams in relation to emotional stability. *J. abnorm. soc. Psychol.*, 1933, 28, 194–197.

10. BLAKE, H., & GERARD, R. W. Brain potentials during sleep. *Amer. J. Physiol.*, 1937, 119, 692–703.

11. BLANCHARD, P. A study of subject matter and motivation of children's dreams. *J. abnorm. soc. Psychol.*, 1926, 21, 24–37.

12. BOLLI, L. Le rêve et les aveugles. I. Le rêve et les aveugles-nés. *J. Psychol.*, 1932, 29, 20–73.

13. CALKINS, M. W. Statistics of dreams. *Amer. J. Psychol.*, 1893, 5, 311–343.

14. CASON, H. The nightmare dream. *Psychol. Monogr.*, 1935, 46, No. 5 (Whole No. 209).

15. CHRYSANTHIS, K. The length and depth of sleep. *Acta med. Orient.* (Jerusalem), 1945, 5, 152–155.

16. CLAVIÈRE, J. La rapidité de la pensée dans le rêve. *Rev. Phil.*, 1897, 43, 507–509.

17. CUBBERLY, A. J. The effect of tensions of body surface upon normal dream. *Brit. J. Psychol.*, 1923, 13, 243–265.

18. DANIELS, G. E. An approach to psychological control studies of urinary sex hormones. *Amer. J. Psychiat.*, 1943, 100, 231–239.

19. DAVIS, H., DAVIS, P. A., LOOMIS, A. L., HARVEY, E. N., & HOBART, G. Human brain potentials during the onset of sleep. *J. Neurophysiol.*, 1938, 1, 24–38.

20. DESANCTIS, S. Sogni nei delinquenti. *Arch. di Psichiat.*, 1896, 5–6.

21. DESANCTIS, S. Emozioni e sogni. *Driter. Int. Congr. für Psychol.* Vol. 44. Munich: Lehman, 1897. Pp. 348–384.

22. DESANCTIS, S. Experimental investigation of dreaming. *Psychol. Rev.*, 1902, 9, 254–282.

23. DESPERT, J. L. Dreams in children of pre-school age. In Freud, Anna, Hartman, H., & Kris, E. (Eds.), *The psychoanalytic study of the child.* Vol. 3 & 4. New York: International Universities Press, 1949. Pp. 141–180.

24. DEUTSCH, E. The dream imagery of the blind. *Psychoanal. Rev.*, 1928, 15, 288–293.

25. EGGER, V. La durée apparente des rêves. *Rev. Phil.*, 1895, 40, 41–59.

26. EISINGER, K., & SCHILDER, P. Dreams and labyrinthine lesions. *Mschr. Psychiat. Neurol.*, 1929, 73, 314–329.

27. ELLIS, H. *The world of dreams.* New York: Houghton Mifflin. 1922.

28. ERICKSON, M. H. On the possible occurrence of a dream in an eight-month-old infant. *Psychoanal. Quart.*, 1941, 10, 382–384.

29. FARBER, L. H., & FISHER, C. An experimental approach to dream psychology through use of hypnosis.

Psychoanal. Quart., 1943, **12**, 202–216.

30. FINLEY, C. S. Endocrine stimulation as affecting dream content. *Arch. Neural. Psychiat.*, 1921, **5**, 177–181.

31. FOSTER, J. C., & ANDERSON, J. E. Unpleasant dreams in childhood. *Child Develpm.*, 1936, **7**, 77–84.

32. FOUCAULT, M. *Le rêve, études et observations.* Paris: Alcan, 1906.

33. FRANKLIN, J. C., SCHIELE, B. C., BROŽEK, J., & KEYS, A. Observations of human behavior in experimental semistarvation and rehabilitation. *J. clin. Psychol.*, 1948, **4**, 28–45.

34. FREUD, S. *The interpretation of dreams.* New York: Macmillan, 1939.

35. FROST, E. P. Dreams. *Psychol. Bull.*, 1915, **12**, 22–25.

36. FROST, E. P. Dreams. *Psychol. Bull.*, 1918, **15**, 12–15.

37. GAHAGAN, L. Sex differences in recall of stereotyped dreams, sleep-talking, and sleep-walking. *J. genet. Psychol.*, 1936, **48**, 227–236.

38. GÖTTKE, L. Über das Traumleben der Epileptiker. *Arch. f. Psychiat.*, 1934, **101**, 136–163.

39. GRIFFITHS, R. *A study of imagination in early childhood.* London: Kegan Paul, Trench, Trubner, & Co., 1935.

40. GRINKER, R. R., & SPIEGEL, J. *Men under stress.* Philadelphia: Blakiston, 1945.

41. GRÜNBAUM, A. M. Die Erforschung der Träume als eine Methode der topischen Diagnostik bei Grosshirnkrankungen. *Z. Neurol. Psychiat.*, 1933, **93**, 416–420.

42. HACKER, F. Systematische Traumbeobachtungen mit besonderer Berücksichtigung der Gedanken. *Arch. ges. Psychol.*, 1911, **21**, 1–131.

43. HALL, C. S. Frequencies in certain categories of manifest content and their stability in a long dream series. *Amer. Psychologist*, 1948, **3**, 274. (Abstract)

44. HALL, C. S. What people dream about. *Scient. American*, 1951, **184**, 60–63.

45. HAMILTON, G. V. *A research in marriage.* New York: Lear, 1948.

46. HANKS, L. M., JR. An exploration of the content of dreams through an interpretation of dreams of convicts. *J. gen. Psychol.*, 1940, **23**, 31–46.

47. HEERMANN, G. Beobachtungen und Betrachtungen über die Träume der Blinden. *Mschr. für Med.* (Leipzig), 1838, **3**, 116–180.

48. HEERWAGEN, F. Statische Untersuchungen über Träume und Schlaf. In Wundt's *Phil. Stud.*, 1888, **2**, 88.

49. HILDEBRANDT, F. W. *Der Traum und seine Verwertung fürs Leben.* Leipzig: Schloemp, 1875.

50. HOFF, A., & PÖTZL, O. Über die labyrintharen Beziehungen von Flugsensation und Flugträumen. *Mschr. Psychiat. Neurol.*, 1937, **97**, 193–211.

51. HOLMBERG, A. R. Nomads of the long bow: the Siriono of Eastern Bolivia. *The Smithsonian Inst. Soc. Anthrop. Publ. No. 10.* Washington, D. C., 1950.

52. HUSBAND, R. W. Sex differences in dream contents. *J. abnorm. soc. Psychol.*, 1936, **30**, 513–521.

53. IKIN, A. G., PEAR, T. H., & THOULESS, R. H. The psychogalvanic phenomenon in dream analysis. *Brit. J. Psychol.*, 1924, **15**, 23–44.

54. JASTROW, J. Dreams of the blind. *New Princeton Rev.*, 1888, **5**, 18–34.

55. JELLIFFE, S. E. Two morphine color dreams with a note on the etiology of the opium habit. *Psychoanal. Rev.*, 1944, **31**, 128–132.

56. JERSILD, A. T., MARKEY, F. V., & JERSILD, C. L. Children's fears, dreams, wishes, daydreams, likes, dislikes, pleasant and unpleasant memories. *Child Devel. Monogr.* No. 12. New York: Teachers Coll., Columbia Univer., 1933.

57. KANT, O. Dreams of schizophrenic patients. *J. nerv. ment. Dis.*, 1942, **95**, 335–347.

58. KARDINER, A. *The traumatic neuroses of war.* New York: Hoeber, 1941. Pp. 88–95.

59. KEYS, A., BROŽEK, J., HENSCHEL, A., MICKELSON, O., & TAYLOR, H. L. *The biology of human starvation.* Minneapolis: Univer. of Minnesota Press, 1950.

60. KIMMINS, C. W. *Children's dreams.* London: George Allen and Unwin, 1937.

61. KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male.* Philadelphia: Saunders, 1948.

62. KLEIN, D. B. The experimental production of dreams during hypnosis. *Univer. Texas Bull.* No. 3009, March, 1930.

63. KLEITMAN, N., MULLIN, F. J., COOPERMAN, N. R., & TITELBAUM, S. *Sleep characteristics.* Chicago: Univer. of Chicago Press, 1937.

64. LADD, G. T. Contribution to the psychology of visual dreams. *Mind*, 1892,

2, 299–304.

65. LOOMIS, A. L., HARVEY, E. N., & HOBART, G. A. Electric potentials in the human brain. *J. exp. Psychol.*, 1936, 19, 249–279.

66. LOOMIS, A. L., HARVEY, E. N., & HOBART, G. A. Cerebral states during sleep as studied by human brain potentials. *J. exp. Psychol.*, 1937, 21, 127–144.

67. LOOMIS, A. L., HARVEY, E. N., & HOBART, G. A. Distribution of disturbance patterns in the human electroencephalogram, with special reference to sleep. *J. Neurophysiol.*, 1938, 1, 413–430.

68. LOVETT DOUST, J. W. Studies of the physiology of awareness: the incidence and content of dream patterns and their relationship to anoxia. *J. ment. Sci.*, 1951, 97, 801–811.

69. MCFARLAND, R. A. Psychophysical studies at high altitudes in the Andes. *J. comp. Psychol.*, 1937, 24, 147–188.

70. MCGLADE, H. B. Relationship between gastric motility, muscular twitching, during sleep and dreaming. *Amer. J. Digest. Dis.*, 1942, 9, 137–140.

71. MACWILLIAM, J. A. Blood pressure and heart action in sleep and dreams: their relationship to haemorrhages, angina, and sudden death. *Brit. med. J.*, 1923, 2, 1196–1200.

72. MALAMUD, W., & LINDER, F. E. Dreams and their relationship to recent impressions. *Arch. Neurol. Psychiat.*, 1931, 25, 1081–1099.

73. MANACÉINE, MARIE DE. *Sleep: its physiology, pathology, hygiene, and psychology.* London: Walter Scott, 1897.

74. MAURY, A. *Le sommeil et les rêves.* Paris: Didier, 1861.

75. MAX, L. W. An experimental study of the motor theory of consciousness. III. Action-current responses in deafmutes during sleep, sensory stimulation and dreams. *J. comp. Psychol.*, 1935, 19, 469–486.

76. MIDDLETON, W. C. Nocturnal dreams. *Scient. Mon.*, 1933, 37, 460–464.

77. MIDDLETON, W. C. The frequency with which a group of unselected college students experience color dreaming and color hearing. *J. gen. Psychol.*, 1942, 27, 221–229.

78. MILES, W. R. The sex expression of men living on a lowered nutritional level. *J. nerv. ment. Dis.*, 1919, 49, 208–224.

79. MONGE, C. *Les erythrémies de l'altitude.* Paris: Masson et cie., 1929.

80. MONROE, W. S. Note on dreams. *Amer. J. Psychol.*, 1897, 9, 413–414.

81. MONROE, W. S. A study of taste dreams. *Amer. J. Psychol.*, 1899, 10, 326–327.

82. MONROE, W. S. Mental elements of dreams. *J. Phil. Psychol. Sci. Math.*, 1905, 2, 650–652.

83. MOWRER, O. H., & MOWRER, WILLIE MAE. Enuresis—a method for its study and treatment. *Amer. J. Orthopsychiat.*, 1938, 8, 436–459.

84. NACHMANSOHN, M. Concerning experimentally produced dreams. In D. Rapaport (Ed.), *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951. Pp. 256–287.

85. NOBLE, D. A study of dreams in schizophrenia and allied states. *Amer. J. Psychiat.*, 1951, 107, 612–616.

86. PIÉRON, H. La rapidité des processus psychiques. *Rev. Phil.*, 1903, 55, 89–95.

87. PINTNER, R., & LEV, J. Worries of school children. *J. genet. Psychol.*, 1940, 56, 67–76.

88. POETZL, O. Experimentell Erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Z. Neurol. Psychiat.*, 1917, 37, 278–349.

89. RICHARDS, O. W. The dream literature. *Psychol. Bull.*, 1924, 21, 338–346.

90. ROFFENSTEIN, G. Experimentelle Symbolträume: ein Beitrag zur Diskussion über Psychoanalyse. *Z. Neurol. Psychiat.*, 1924, 87, 362–372.

91. ROWLAND, L. W. The relation of judgments of excitement value to certain bodily changes shown during hypnotic dreams. *Proc. Okla. Acad. Sci.*, 1932, 12, 92. (Abstract)

92. SARASON, S. B. Dreams and Thematic Apperception Test stories. *J. abnorm. soc. Psychol.*, 1944, 39, 486–492.

93. SCANTLEBURY, R. E., & PATTERSON, T. L. The effect of certain psychic phenomena on various phases of gastric hunger contractions. *Proc. Amer. J. Physiol.*, 1938, 123, 179–180. (Abstract)

94. SCANTLEBURY, R. E., FRICK, H. L., & PATTERSON, T. L. The effect of normal and hypnotically induced dreams on gastric hunger movements of man. *J. appl. Psychol.*, 1942, 26, 682–691.

95. SCHRIEVER, W. Einige Traumbeobachtungen. *Z. Psychol.*, 1935, 134, 349–371.

96. SCHRÖTTER, K. Experimentelle Träume. *Zbl. Psychoanal.*, 1912, 2, 638–646.

97. Schroetter, K. Experimental dreams. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951. Pp. 234–256.

98. Schubert, H. J. P., & Wagner, M. E. The relation of individual personal data responses and transiency, place among siblings, and academic ability. *J. abnorm. soc. Psychol.*, 1936, **30**, 474–483.

99. Selling, L. S. Effect of conscious wish upon dream content. *J. abnorm. soc. Psychol.*, 1932, **27**, 172–178.

100. Sirna, A. A. An electroencephalographic study of the hypnotic dream. *J. Psychol.*, 1945, **20**, 109–113.

101. Small, M. Dreams. *Psychol. Bull.*, 1920, **17**, 346–349.

102. Sorokin, P. A. *Man and society in calamity*. New York: Dutton, 1942.

103. Stepanow, G. *Sogni Indottistudio sperimentale sull'influenze degli stimoli acustici sul sogno*. Florence: Aldino, 1915.

104. Straecke, J. Neue Traumexperimente im Zusammenhang mit aelteren und neueren Traumtheorien. *Jb. Psychoanal. Psychopath. Forsch.*, 1913, **5**, 233–306.

105. Sweetland, A., & Quay, H. An experimental investigation of hypnotic dream. *J. abnorm. soc. Psychol.*, 1952, **47**, 678–682.

106. Thomson, E. R. An inquiry into some questions connected with imagery in dreams. *Brit. J. Psychol.*, 1914, **7**, 300–315.

107. Titchener, E. B. Taste dreams. *Amer. J. Psychol.*, 1895, **6**, 505–508.

108. Tobolowska, J. *Etude sur les illusions du temps dans les rêves du sommeil normal*. Thèse de Paris, 1900.

109. Trapp, C. E., & Lyons, R. H. Dream studies in hallucinated patients; preliminary study. *Psychiat. Quart.*, 1937, **11**, 515–538.

110. Vaschide, N. Les recherches expérimentelles sur les rêves. *Rev. Psychiat.*, 1902, **5**, 145–164.

111. Vold, J. M. Einige Experimente über Gesichtsbilder im Träume. *Z. Psychol.*, 1896, **13**, 66–74; or *Dritter Int. Congr. für Psychol.* (Munich), 1897, 355–357.

112. Vold, J. M. *Über den Traum*. O. Klem (Ed.). Leipzig, 1910.

113. Wada, T. An experimental study of hunger in its relation to activity. *Arch. Psychol.*, 1922, **8**, No. 57.

114. Walsh, W. S. Dreams of feebleminded. *Med. Rec.*, 1920, **97**, 395–398.

115. Weed, Sarah, & Hallam, Florence. A study of dream consciousness. *Amer. J. Psychol.*, 1896, **7**, 405–411.

116. Welch, L. The space and time of induced hypnotic dreams. *J. Psychol.*, 1936, **1**, 171–178.

117. Wheeler, R. H. Visual phenomena in the dreams of a blind subject. *Psychol. Rev.*, 1920, **27**, 315–322.

118. Wiggam, A. A contribution to the data of dreaming. *Ped. Sem.*, 1909, **16**, 240–251.

119. Wile, I. S. Auto-suggested dreams as a factor in therapy. *Amer. J. Orthopsychiat.*, 1934, **4**, 449–463.

120. Witty, P. A., & Kopel, D. The dreams and wishes of elementary-school children. *J. educ. Psychol.*, 1939, **30**, 199–205.

121. Woodworth, R. S. Note on the rapidity of dreams. *Psychol. Rev.*, 1897, **4**, 524–526.

ASSESSING SIMILARITY BETWEEN PROFILES[1]

LEE J. CRONBACH   AND   GOLDINE C. GLESER
University of Illinois   *Washington University School of Medicine*

A great many current investigations, particularly in clinical and social psychology, deal with similarity between profiles of test scores. Such studies vary widely with regard to the problems posed and the specific variables used, but they have in common an attempt to deal with several scores or traits simultaneously. Some investigators attempt to identify "types" of people who have similar configurations of scores. Much of so-called inverse factor analysis has this aim. Other studies attempt to differentiate clinical or occupational groups by means of patterns of test scores (e.g., 1, 28). In another type of problem, two or more profiles for the same person are compared. The person is assessed more than once on the same set of variables, and the consistency of the profiles is measured. This is one method used to study the validity of clinical procedures (5, 24). Profile comparison also permits exploration of new variables such as self-consistency over time (31) and assumed similarity in perception of others (17).

At present many techniques are available to the investigator who is concerned with assessing the degree of profile similarity. The method most widely known among psychologists is that of correlating one profile with another, generally termed a Q correlation. Burt (3) and Stephenson (32) have been chiefly responsible for developing this approach.[2] Special indices have also been proposed, such as the coefficients of pattern similarity of Cattell and those of du Mas. A distance measure has been described recently by Osgood and Suci

(26) and by the present writers (11).[3]

A very valuable summary of statistical literature bearing on the use of profiles or patterns to classify individuals into relatively homogeneous groups has been prepared by Hodges (22). Other recent reviews which deal in part with this problem are Gaier and Lee's (20) and Tyler's (35).

The various available methods of measuring profile similarity yield somewhat different results. Proper choice of a measure for a specific investigation requires knowledge of the assumptions, limitations, and information utilized in the several methods of measuring profile similarity. *It appears that the methods most often used have serious limitations. Much superior methods can be proposed.*

We intend in this paper to examine in a general way the problem of comparing sets of scores and to clarify the mathematical logic involved

[2] Stephenson's current work on Q technique (33) departs from the correlational methods reviewed here. We shall not discuss here the logic of his basically new approach using analysis of variance.

[3] The work of Osgood and Suci (26), and our own work, was in large measure independent. While working on our separate problems, however, we exchanged ideas occasionally, and found our interests converging on the D measure. We appreciate their cooperation and that of others who have discussed our problem with us.

therein. This permits us to consider the various formulas which have been advanced in the past, and to draw attention to those approaches which seem to have greatest merit.

This paper is primarily concerned with *descriptive* indices applicable to the investigation of questions such as the following:

1. How similar are Persons 1 and 2?
2. How similar is Person 1 to Group *Y?*
3. How homogeneous are the members of Group *Y?*
4. How similar is Group *Y* to Group *Z?*
5. How much more homogeneous is Group *Y* than Group *Z?* Than the combined sample?

Comparable questions may be asked in studies concerned with two or more profiles for the same person.

While it is necessary to describe the degree of similarity between score sets in many of the investigations now being pursued, it is often equally or more important to test hypotheses such as "Group *Y* and Group *Z* can be regarded as samples from the same population" or "Individual 1 is more likely to be a member of Group *Y* than of Group *Z*." Such problems of *inferential* statistics relevant to multivariate analysis have been thoroughly studied by Fisher, Hotelling, and the Calcutta school, and several significance tests are available for normally distributed variables (29). We shall not discuss the inferential problems, being concerned solely with descriptive formulas for reporting degree of similarity.

General Methodological Difficulties

While the procedures permitted to the investigator of profile similarity are varied, they involve numerous pitfalls. We shall discuss some of these difficulties as a preliminary to formal analysis of profile comparison methods.

Similarity as a general quality. Thinking of persons as "similar" or "dissimilar" is a common oversimplification. This attractive notion, however, does violence to a fundamental principle. If behavior is described in terms of independent dimensions, then persons who are similar in one dimension may be no more similar in some second dimension than persons who are dissimilar in the first dimension. *In other words, similarity is not a general quality. It is possible to discuss similarity only with respect to specified dimensions (or complex characteristics).* This means that the investigator who finds that people are similar in some set of scores cannot assume that they are similar in general. He could begin to discuss general similarity only if his original measurement covered all or a large proportion of the significant dimensions of personality. Thus any problem inquiring whether similar people perform differently from dissimilar people must be stated in terms of the question "Similar in what?" It is most unlikely that similarity in every quality has the same effect.

Reduction of the configuration by similarity indices. Many investigators are attracted to profile similarity studies because they believe that in this way they can take into account the entire configuration of scores. However, when we try to treat a set of scores by any of the mathematical methods now being used, we no longer study the entire configuration. Instead, by reducing the configuration or the relationship between two configurations to a single index, we discard much of the information in the score set.

We may illustrate this by referring to Gage's study of insight (19). He asked a teacher to predict the responses of a pupil. He scored the

predictions using the responses actually given by the pupil as a key, thus estimating the accuracy of the teacher's perception. It is obvious, however, that a more refined question could be asked regarding the teacher's ability to perceive separate aspects of the pupil. In using a total index, Gage was forced to combine these many separate aspects of insight into an over-all score. It is important that the investigator recognize the limitations of so-called global approaches even though they may be the best for him to use in initial exploration of a particular area.

Absolute interpretation of index. Another type of difficulty which frequently complicates interpretation of profile similarity studies is the failure to recognize that the magnitude of the similarity index has no meaning in itself. In conventional psychometrics, we would not give serious attention to the absolute value of a test score. When we compare a person to a key, the number of items on which he and the key agree is a form of correlation. We are aware that we should not interpret this raw score which reflects the difficulty of the items. Instead, we give our attention to the relative standing of the individual in some reference group. Correlations between persons, and other similarity indices, entail precisely the same problem. Too often, accustomed to interpreting correlations as absolute numbers, investigators interpret similarity indices without recognizing that they also depend upon the difficulty or popularity of the items or tests.

One often cited study by Fosberg shows the fallacy of this type of interpretation (18). Fosberg hoped to demonstrate that the Rorschach test is proof against faking. He therefore asked individuals to take the test in the normal manner, and then to take

it attempting to make the best possible impression. He correlated the two psychograms for a given individual and interpreted the resulting high correlations as showing that the Rorschach was proof against faking. Now it is true that the psychogram under "fake good" conditions could be predicted from that under normal conditions. But the "fake good" psychogram could have been predicted quite well from the psychogram of some other person chosen at random. Between any two Rorschach records taken at random, there will tend to be a high correlation just because certain scores (e.g., D, F) will usually be large, and other scores (e.g., m, cF) will usually be small.

It is evident that any estimate of the similarity of particular profiles must be evaluated relative to the similarity of people in general on the measures in question. A high index of similarity between two persons might indicate that they are unusually alike, or might indicate that they possess in common only the characteristics most humans have. For example, Gage (19) considered insight to be indicated by a marked similarity between prediction and actual response. He found that a large part of the correlation between predicted response and actual response was accounted for by the teacher's ability to predict the responses of pupils in general. When the teacher was asked to predict the *average* response of pupils, the correlation with the actual responses of an individual pupil was frequently as high as when the teacher attempted to predict that particular pupil's response.

Noncomparability of scale units. Combining many traits into any sort of composite index, whether it be a D measure, a Q correlation, a discriminant function, or any of the

other methods presently used, involves assumptions regarding the scale of measurement which usually cannot be defended (7, 8). If, for example, one score measures intelligence and a second one reflects anxiety level, any single index based on this profile involves an assumption that one unit of intelligence is equivalent to some number of units on the anxiety scale. Such an assumption is perhaps necessary if it permits investigations which would otherwise be impossible. It may also be possible to justify the units assigned to the respective scales by a mathematical treatment which selects the weights to maximize some prediction. This is an empirical solution, however, and does not contribute directly to development of theory.

A GENERALIZED CONCEPT OF PATTERN SIMILARITY

We now introduce a model for the concept of similarity between persons which provides a basis for systematic discussion of the assumptions underlying most of the common measures of profile similarity.

A profile or pattern pertaining to a person consists of a set of scores. We shall use the following notation:

j = any of the variates a, b, c, $\cdots$ which are k in number;

i = any one of the persons 1, 2, $\cdots$ N;

x_{ji} = the score of person i on variate j.

Considering only two persons, we have the set of $x_{j1}(x_{a1}, x_{b1}, \cdots x_{k1})$ for person 1, and the set of x_{j2} for person 2. Without placing any restriction upon our data, we may regard the x_{j1} as the coordinates of a point P_1 in k-dimensional space. The x_{j2} define a point P_2. The more similar the measures of two individuals the closer will their points lie

in the k-dimensional space, and, conversely, the further apart the points the more dissimilar are the corresponding measurements. Accordingly we define the *dissimilarity* of two individuals as the linear distance between their respective points.

If we represent the variables by orthogonal axes, the distance D between any two points may be easily obtained by use of the generalized Pythagorean rule,

$$D_{12}{}^2 = \sum_{j=1}^{k} (x_{j1} - x_{j2})^2. \quad [1]$$

D^2 can be used directly as a measure of similarity. In most cases, however, it is preferable to obtain D, since the larger differences between persons are much exaggerated in squaring. D is less skewed than D^2 but is not normally distributed.

Formula [1] is a general expression for the dissimilarity between two profiles. It may be applied to practically any type of score set; viz., responses to a series of items, raw scores on a set of tests, profiles of deviation scores, ratings of a group of stimuli on a subjective scale, or responses in a Stephenson forced-sort procedure. While formula [1] results in a measure of dissimilarity no matter what types of scores are used, the interpretation of the results depends on the nature of the scores.

One basic decision made by the investigator is whether to work with the original score set or to convert it by centering about the person's mean or by standardizing within the person. He may make these conversions before using formula [1]. Many of the current formulas automatically introduce some such treatment of scores. Converted scores in general alter the domain within which similarity is measured and consequently alter the results.

Elevation and scatter within profiles. A set of k scores, whether expressed in raw or standard measure, has k degrees of freedom and may be considered as a configuration in k space. When the profile is expressed as a set of deviations about the person's mean or when the profile is standardized within the person, the number of degrees of freedom is reduced. This has important consequences. In order to discuss them we introduce the terms *elevation, scatter,* and *shape.* *Elevation* is the mean of all scores for a given person. *Scatter* is the square root of the sum of squares of the individual's deviation scores about his own mean; that is, it is the standard deviation within the profile, multiplied by $\sqrt{k}$. *Shape* is the residual information in the score set after equating profiles for both elevation and scatter. We can clarify these terms by introducing numerical illustrations. Suppose that we have five traits a, b, c, d, e, and persons A, B, and C.

	a	b	c	d	e
A	2	-2	0	3	2
B	0	-4	-2	1	0
C	3	-1	3	-1	-4

According to formula [1], D_{AB}^2 is 20. $D_{AC}^2 = D_{BC}^2 = 63$.

Elevation is determined by averaging the scores for each individual. For the example above, the elevations are as follows: A, 1; B, -1; C, 0. Removing elevation, the individual profiles become:

	a	b	c	d	e
A	1	-3	-1	2	1
B	1	-3	-1	2	1
C	3	-1	3	-1	-4

Now the distance between A and B is 0. Those persons who are different when their total profiles are taken into account are indistinguishable on the basis of their profiles of deviation scores. D_{AC}^2 and D_{BC}^2 now equal 58.

The operation of eliminating differences in elevation from the profiles is referred to by Thomson (34) and others as centering about persons. Geometrically, it is equivalent to projecting all persons into a $k-1$ space orthogonal to the line defined by the equations $a = b = c \cdots$. Comparison of deviation scores is involved in testing certain hypotheses regarding scatter in mental tests (2). Burt eliminates elevation when he obtains a matrix of covariances between profiles for use in factoring persons into types (3). If we use D' as a symbol for distance between profiles after projection into $k-1$ space, we have the following equation:

$$D'_{12}{}^2 = D_{12}{}^2 - k\Delta^2 El_{12} \qquad [2]$$

Here ΔEl represents the difference in elevation between the two persons. It is evident that the difference between persons has two components, one due to elevation, and one due to the remaining information in the profile. Treatment of deviation scores discards information about differences in elevation.

When differences in scatter between profiles are eliminated, the measure of similarity is reduced to a consideration of shape alone. This is accomplished by dividing each deviation score by the individual's scatter, thus standardizing the profile. Geometrically, this operation amounts to projecting every score set in $k-1$ space onto a $k-2$ hypersphere. The center of the hypersphere is at the point representing in $k-1$ space a completely flat profile.

If for each of the three persons in the above example we divide his deviation profile by his scatter, we obtain the following new profiles:

	a	b	c	d	e
A	1/4	−3/4	−1/4	1/2	1/4
B	1/4	−3/4	−1/4	1/2	1/4
C	1/2	−1/6	1/2	−1/6	−2/3

Now $D_{AC}^2 = D_{BC}^2 = 2.25$. $D_{AB}^2 = 0$ as before.

Letting D'' be our symbol for the D measure obtained from two standardized score sets,

$$D''^2 = \frac{D'^2 - \Delta^2 S}{S_1 S_2} = \frac{D^2 - k\Delta^2 El - \Delta^2 S}{S_1 S_2}. \quad [3]$$

Here S is the scatter of an individual and ΔS is the difference in scatter. It is clear from this equation that by standardizing the profile we eliminate from consideration one further type of difference between the persons.

Elevation and scatter have commonly been eliminated in past studies of similarity between persons. It is easily shown that

$$D''^2 = 2(1 - Q) \quad [4]$$

where Q is the product-moment correlation between scores. It will be recalled that in product-moment correlation, one subtracts the product of means from the cross-product terms, and divides by the standard deviations (which are proportional to the measures of scatter). In other words, all correlations between profiles are essentially measures of distance in $k - 2$ space.

Equations [3] and [4] make clear that D in k space will, in general, not give the same result as Q for a given pair of score sets, nor can D be inferred from factor loadings derived from Q correlations. Osgood and Suci (26) demonstrated close correspondence between the two sorts of measures, but only for an unusual set of data where ΔEl and ΔS are small. Warrington (36) determined the extent to which information is discarded in various treatments by building hypothetical data from a mathematical model. For his analysis, he employed five factors, represented with varied loadings in 60 items. Each of his hypothetical persons was assigned scores, distances between persons were determined, and these were correlated with distance measures based on the factor scores. For perfectly reliable items the similarity measures correlated .92 with the criterion. This "validity" dropped to .85 when elevation was removed from the similarity index but not from the criterion. For items of moderate reliability, the validity dropped from .81 to .55 when elevation was removed.

RELATION OF OTHER FORMULAS TO THE D MEASURE

Table 1 lists the formulas most frequently used in psychological investigations of profile similarity, together with some of their more prominent characteristics.

Treatments in k *space.* The D measure presented in formula [1] considers all k dimensions in the original data. This measure has recently been discussed by Osgood and Suci (26), but a quite similar formula appeared in the literature much earlier, as Pearson's "coefficient of racial likeness" (CRL) (27), which was developed to measure the similarity between two groups or the similarity of an individual to a group. In its original form, CRL was essentially the same as D^2 save that all variates were expressed in standard measure and a multiplier involving the number of cases per group was included.

The Pearson index proved unsatisfactory in the anthropological research for which it was developed. Some of the criticisms arise out of its insensitivity to differences of number of cases from group to group. These

criticisms are irrelevant to our present purpose. Morand (see Rao, **30**) notes that in some anthropological research the index has given unreasonable results for groups which were regarded as quite dissimilar, intuitively or theoretically. From the context, we judge that this difficulty is a consequence of the high weight CRL assigns to general factors among the variates. This problem may arise in measures of similarity whenever variates are intercorrelated. We shall discuss the problem of correlation in more detail later.

Cattell (**6**) has proposed an index r_p which is like D in many respects. He introduces a transformation, however, which makes the obtained index range from 1 to −1. In our notation,

$$r_p = \frac{K\Sigma\sigma_j^2 - kD^2}{K\Sigma\sigma_j^2 + kD^2} \quad [5]$$

where K represents twice the median χ^2 corresponding to the given number of variates. D or r_p would give the same results so far as the *ordering* of dissimilarity is concerned.

Cattell arrived at this index because he believes that similarity should be measured by an index which is comparable to a correlation. This assumption seems to us neither necessary nor desirable. If persons fall into a multivariate normal distribution there should be very many similar pairs, and relatively few pairs who are far from each other. Furthermore, if we are dealing with variates having an unlimited range then no matter how far apart person 1 is from person 2, there is no theoretical reason why there cannot be a person 3 such that $P_1P_3 > P_1P_2$. Therefore we see no reason why the measure of separation should have a limit. "Complete dissimilarity of persons" is an undefinable concept.

Webster (**37**) proposed that intraclass r might have advantages for measuring similarity in k space.

$$r_{in} = 1 - \frac{D_{12}^2}{S_1^2 + S_2^2 + \frac{1}{2}k\Delta^2 El} \quad [6]$$

The denominator in [6] is the sum of squares of scores of both persons about the grand mean of their scores. The larger this denominator the closer will r_{in} approach +1 for pairs having the same D. To illustrate, con-

TABLE 1
SIMILARITY FORMULAS AND THEIR CHARACTERISTICS

Symbol and Proponent	Procedure	Type of Comparison	Remarks
D (Osgood-Suci, **26**; Cronbach-Gleser, **11**)	Distance measure	k (also $k-1$, $k-2$)	A general formula
CRL (Pearson, **27**)	Distance measure for standardized variates	k	
r_p (Cattell, **6**)	Transformed distance measure for standardized variates	k (also $k-1$)	Converts D to a scale from 1 to −1
Q (Stephenson, **32**)	Product-moment correlation across variates	$k-2$	Symbol Q used here instead of r for clarity
Rho (Spearman)	Correlation across scores ranked within a profile	$k-2$	
Tau (Kendall, **25**)	Based on rank arrangements	$k-2$	Highly correlated with rho
r_{ps} (du Mas, **13**)	Based on tally of similarity of slope along profiles	$k-2$	Estimate of tau based on partial data

sider person X with standard scores 1.0, 1.0, on two variates, and person Y with standard scores 1.1, 1.1. For this pair, D is $\sqrt{0.02}$. S for each person is zero, the denominator is small, and r_{in} is -1. In other words, this pair of persons is reported by intraclass r to have maximum dissimilarity, whereas the D measure reports them to be close together. The definition upon which the D measure is based appears to present a more satisfactory conception of similarity than the definition embodied in the intraclass measure.

Treatments in $k-2$ space. Several formulas have the effect of measuring similarity in $k-2$ space. We have already noted that a Q correlation based on raw scores gives the same result as obtaining D from scores standardized within profiles. Correlation is thus a special case of the D measure.

Measures of similarity are at times based on scores ranked from highest to lowest within the profile. The correlation of two such sets of ranks yields rho, which is thus like Q in many of its properties. Sometimes rank correlations are used in the belief that assumptions regarding the test score metric are thereby avoided. This is not the case. When scores are ranked, the separation between two adjacent ranks is fixed over the whole range, forcing all profiles into the same rectangular distribution. This forcing may be justified in certain studies, but it does involve a definite assumption.

Kendall's tau is a rank correlation based on the *direction* of differences between all possible pairs of variates. Tau is very closely related to rho but is somewhat more laborious to compute. In some statistical work, it is an advantage that the sampling distribution for tau is known.

Du Mas has suggested the coeffi-

cient r_{ps}. Kelly and Fiske drew our attention to the fact that r_{ps} is a sort of approximation to tau, in which pairs of adjacent variates only are considered. Results from du Mas' method therefore depend upon the order in which variates are listed in a profile. Different results would be obtained if some other order were used. r_{ps} is biased when the arrangement of traits is not strictly random. Furthermore, it uses relatively little of the information in the profile, and is therefore inexact. r_{ps} does not appear to have advantages over rho or D''.

Should differences in elevation be disregarded? A basic question is whether similarity between score sets is more meaningfully investigated by allowing differences in elevation to affect the result.

Cattell (6) and du Mas (14) have argued that differences in level between profiles are generally important and should be included in the index. For many studies, it is surely desirable not to regard two people as similar if their profiles have the same shape but differ in elevation. In the Wechsler test, for example, the elevation, being the sum of the scores, is a measure of over-all ability. The interpretation of the profile shape is dependent upon elevation. The fact that Vocabulary is higher than Digit Span means something qualitatively different for a college graduate with an IQ of 120 from what it means for a 10-year-old with an IQ of 100. To reduce the data by leaving elevation out of account may cause people to appear similar who are quite different in the domain the investigator desires to study.

On the other hand, there may be studies in which the elevation component is of no interest. If, for example, data are obtained from a personality questionnaire in which a person re-

sponds *yes* or *no* to each item, and the total score in each category is the number of questions marked *yes*, the differences in elevation between persons will be due partly to a response set (9). The investigator may decide that this "*yes*-saying tendency" is irrelevant to his problem, and if so, he will want to eliminate that component from his data. If he makes such a decision, reduction of the data to $k-1$ space is appropriate.

The elevation component in a profile represents the sum (or average) of all scores, and depends on the direction of scoring of the variates. A trait could be scored as "submission," for instance, instead of "dominance"; any such reversal alters the composition of the elevation score. If there is no particular reason for scoring each variate in one direction rather than the other—and this is generally the case unless variates are systematically correlated—then the elevation component is determined arbitrarily by these scoring decisions. It is highly undesirable to eliminate the elevation component when it is thus arbitrarily defined.

If a general factor is present in the variates it is often possible to choose a direction for scoring each variate which yields consistently positive intercorrelations among variates. The elevation factor, when this set of scores is used, will be heavily loaded with the first principal component of the scores, i.e., the general factor (23). This first factor may be an important one to consider in judging the similarity of profiles.

In general, it appears undesirable to eliminate elevation unless the investigator can interpret it definitely as representing individual differences in a quality which he does not wish to take into account in his similarity measure. If he is uncertain as to which is the more appropriate direc-

tion for scoring each of the variates, then the investigator should use the measure D in k space. Ebel (15), working on the problem of similarity of score sets as it is encountered in studying the reliability of ratings, makes a similar recommendation. In his problem, the mean level of ratings assigned by each rater is comparable to our elevation. He lists practical considerations which make it wise at some times, and unwise at others, to consider differences in level in assessing the agreement of raters.

Should differences in scatter be disregarded? Any treatment which equalizes scatter of profiles before computing the difference measure is equivalent to projecting points onto the surface of a hypersphere within the $k-1$ space. This has the effect of increasing the jaggedness of profiles which are relatively flat, or, we might say, of reducing the jaggedness of profiles having a large amount of scatter. This introduces a serious difficulty. Figure 1 illustrates the fact that in projection onto the sphere differences between persons near the center are much magnified. The small D'_{12} becomes a large D''_{12}. D''_{34}, however, is little greater than D'_{34}. Points 1 and 2, near the center of the sphere, represent persons with flat profiles. Persons who would be judged quite similar in k or $k-1$ spaces are sometimes reported as markedly dissimilar in the $k-2$ measure.

Another aspect of the same problem is illustrated in Fig. 2. Any profile contains some error of measurement so that the location of the individual in $k-1$ space is only approximate. We indicate the possible positions in $k-1$ space of each individual over many trials by a cloud of points within the circle. The possible positions a person might take in $k-2$ space are then indicated by
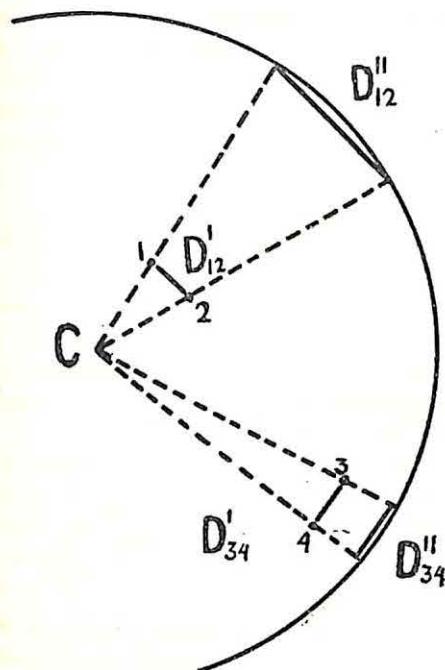
FIG. 1. MAGNIFICATION OF DISTANCES IN PROJECTION ONTO SPHERE

the distribution of points on the edge of the circle. It is clear that the greater the error, the greater the dispersion in $k-1$ and $k-2$ spaces. For a person who has a moderate amount of error and whose scatter is low, the projection in $k-2$ space has almost no meaning. On different trials he might fall anywhere in the $k-2$ space, and it is a matter of chance which persons he is similar to in a particular set of data. Either

a high or low value of Q can arise by chance.

Results from analysis of profiles in $k-2$ space are dependable only when scatter is large relative to the error dispersion for the individual. So long as some profiles may be expected to be flat or nearly so, treatments of these profiles in $k-2$ space will be very much influenced by random error. This difficulty is greatest when most of the variance in the k scores comprising the profile is accounted for by a small number of factors. As more factors are represented in the variate set it is less likely that flat profiles will be obtained.

We must question whether the study of profiles in $k-2$ space, or more specifically whether correlation between profiles, is a justifiable line of investigation. This procedure has the disadvantage of removing the elevation factor and in addition tends to magnify error variance. In general, therefore, we would regard treatments in $k-2$ space as inferior to treatment of the data by D or D'.

Such success as Stephenson and his followers have obtained despite these difficulties may be explained by precautions Stephenson has introduced into his design. For one thing, Stephenson has always employed a large number of variates, each one being an item describing some per-

Low scatter, low error   Low scatter, moderate error   High scatter, moderate error
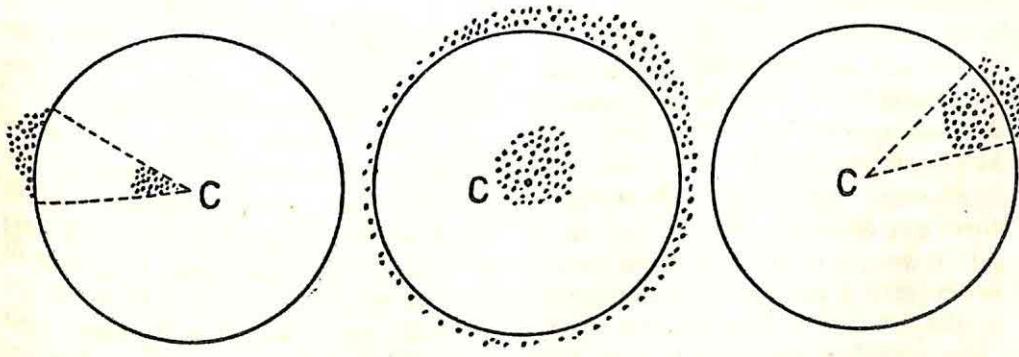
FIG. 2. EFFECT OF ERROR AND SCATTER ON THE PROJECTION ONTO A SPHERE

sonality trait. If the item intercorrelations are not generally positive, the first component removed as an elevation factor is a relatively small proportion of the total variance or information in the profiles. The part removed may be an important portion, but the $k-1$ profile still contains a great deal of useful information. The large number of variates also makes flat profiles in $k-1$ space less frequent.

In Stephenson's "balanced design questionnaire" each item is accompanied by another statement which has approximately the opposite meaning. By this device, Stephenson essentially assures that the sum over all items (i.e., elevation) departs from zero only by chance, and thus no information is eliminated from the data during the statistical elimination of the elevation component.

The magnification of error in projection to $k-2$ space will be slight if few persons have flat profiles. This can be assured by introducing items which have unequal means for the group. Then the centroid of the group will be far from the center of the sphere on which persons are projected. The difficulty with this solution is that, as the centroid of the group moves farther from the center of the sphere, persons are less differentiated in $k-2$ space, and error accounts for a larger proportion of the dispersion.

It is not surprising that most profile studies today utilize comparisons in $k-2$ space, since the problems have been conceived in terms of correlation as used to study relationships between tests. It is questionable, however, whether that model is a particularly good one. In determining the similarity between two *tests*, it is reasonable to eliminate the mean and variance from consideration. As Thomson (34) and Burt (4) have

pointed out, the test mean represents its general level of difficulty for the population, while the variance is a function of the units used. Differences between tests in these values are usually quite arbitrary, depending on the choice and number of items. When we are mainly interested in the underlying relationship between tests these differences are of no importance and are neglected in the correlation formula. In dealing with similarity of *individuals*, however, it is necessary to consider rather carefully what logic is involved when individuals are equated for level and scatter.

Measures in $k-2$ space can give useful information only if both the dispersion of persons in $k-1$ space and the scatter for nearly all persons are large relative to the error dispersion. Data in $k-1$ space are required to determine whether these conditions are met. Then one can determine whether profiles in $k-1$ space are reliable, and whether there are many flat profiles. The investigator can, if he wishes, eliminate the people with flat profiles from the study. The forced-sort does not collect data on scatter, and one has no basis for judging which profiles are reliably located.

It seems quite important for those studying similarity to investigate reliability directly by obtaining two estimates for each profile. Reliability of $k-2$ space measures has ordinarily not been examined in past investigations of similarity.

In those studies where $k-2$ space measures have been used in the past, properly interpreted positive results need not be discounted. The faults to which we have drawn attention operate to obscure true relations and to make the measurement technique insensitive. This would make nonsignificant results likely in some instances where a better technique

would find more relationship. It would tend to make particular Q correlations or differences between such correlations undependable and inconsistent.

In summary, our consideration of all possibilities leads us to the opinion that the most generally advisable procedure for comparing profiles is to employ D in k space, except where it is known that the elevation factor is saturated with a variable which it is desired not to consider.

CONTRIBUTION OF EACH VARIATE TO THE SIMILARITY MEASURE

The Mahalanobis distance. A formula which we have not discussed to this point is the generalized distance measure of Mahalanobis (see Rao, **30**). The Mahalanobis distance is found from the formula:

$$\mathbf{D}^2 = \sum_{j} \sum_{j'} \alpha^{jj'} \Delta x_j \Delta x_{j'} \qquad [7]$$

where $\alpha^{jj'}$ is the jj' element of the inverse of the covariance matrix between variates within groups. We use $\mathbf{D}$ to distinguish this measure from our D. The Mahalanobis measure was designed for the purpose of measuring the distance between groups, rather than between individuals, but the formula can also be interpreted as related to the difference between individuals. If this is attempted, the intercorrelations of the variates for an appropriate reference group must be known.

The D measure is a measure of similarity in which the orthogonal components of the original set of variates are assigned *equal weight*. In other words, the complex formula presented above yields the same results as would be obtained if one factored the correlation matrix into k orthogonal factors, computed the person's scores on these components, and then applied the D formula to

measure similarity. For variates which are standardized and uncorrelated $\mathbf{D}$ is identical to D.

$\mathbf{D}$ has several interesting properties. It has a known distribution function and thus forms a basis for testing the significance of a difference between groups. Moreover, $\mathbf{D}$ is closely related to Fisher's discriminant function, and particularly to the proportion of individuals classified into the wrong group by the most efficient possible discriminant function (**30**, p. 180). It is not, however, especially suited to the descriptive problem which we are discussing.

In any set of correlated variates, some variance is due to general qualities or factors represented in several variates, some due to meaningful factors found only in a single variate, and some due to error of measurement. In a principal-components analysis, k factors will be determined but the last factors may be almost entirely due to error of measurement. The Mahalanobis measure weights unreliable and unimportant factors equally with the first few components in the variates. That is, it assumes that any k variates represent k equally important factors. This is undesirable in a descriptive index, since differences between individuals on factors which are not well represented in the test battery will be unstable from one trial to another, and hence $\mathbf{D}$ for individuals will be unstable. When the formula is applied to differences between groups, no such problem arises, for groups will show negligible differences on factors which consist largely of error.

Weights in the D measure. The interpretation of the D measure is facilitated if we consider what weight it assigns to the orthogonal components underlying the variates. Some investigators have proposed that uncorrelated scores be employed in any

that the investigator may properly use D or D_w, whether variates are correlated or not. He should give careful thought to the question of whether or not to standardize variates. In many studies of similarity it is probably desirable to perform a factor analysis on the matrix of correlations or covariances among tests before studying similarity of persons. This permits the investigator to select his set of traits or their weighting on a more intelligent basis than he could without the factor analysis.

Cluster scoring. It may often be desirable to employ many items to measure a much smaller number of traits. This is the plan used in assembling items for many tests (e.g., Kuder, Guilford-Martin). Consideration should therefore be given to special problems arising for such a set of items. A particularly important question is whether the *items* should be treated as variates in the D measure, or whether scores on *clusters* of items (i.e., subtests) should be used.

When assembling groups of items to measure particular traits it is difficult for the investigator to make sure that these traits will have the desired weight in the D measure based on item scores. The principal components of the items will not be the same as the intended traits. Each trait will be a complex and unknown combination of the principal components. Its weight will depend highly on the choice of items and their particular factor structure.

The investigator has several possible procedures which may help him to approach the desired weights. Stephenson has suggested constructing items which systematically sample the domain of traits under consideration (32). If this sampling were perfect, he would insure uniform coverage of the domain so that the

traits would be uniformly weighted. This approach is likely to succeed only if the item writer has more knowledge of the factorial structure of personality items than is presently available. Another solution is to perform a factor analysis on the set of items, then rotate to the desired factor solution and obtain trait scores on which to compute similarity measures. This, however, is generally impractical.

In some cases a more practical solution is to combine items into groups or clusters and obtain subtest scores for each person. Such cluster scoring is feasible only when there is a logical or statistical basis for combining items. Cluster scoring may be based on a priori grouping of items, but these groupings should be analyzed for internal consistency. From the matrix of intercorrelations of the pool of items, it would be possible to assign items to relatively homogeneous subtests (12).

D based on cluster scores weights the underlying components of the items differently from D based on the original items. In the cluster distance, the element common to the several items is given greater weight than it has when the distances on the separate items are combined. The sum of a group of items gives relatively great weight to factors present in more than one item (10, 23). If specific factors each present in only one item are not especially important, cluster scoring reduces their combined weight in order to give greater weight to the common element running through a whole group of items. To give a specific illustration, a score on hypochondriasis or health adjustment based on a number of items will give great weight to a general tendency to claim somatic symptoms. It will give less weight

than the item scores to specific symptoms such as a tendency to have colds or to have headaches.

In the same manner that cluster scoring reduces the weight given to specifics, it also reduces the weight given to differences between persons arising from error of measurement. Hence cluster scores, and similarity measures based on them, will be more reliable than scores based on the items. Warrington (36), with his hypothetical data, has confirmed this greater dependability of cluster-scored profiles. For one particular criterion, for instance, using Q-sort data, he found these validity coefficients:

D measure based on items as variates, perfect item reliability .70
D measure based on clusters as variates, perfect item reliability .74
D measure based on items as variates, moderate item reliability .18
D measure based on clusters as variates, moderate item reliability .66

It is apparent that cluster scoring overcomes much of the loss of information due to item unreliability. Stephenson is now essentially using cluster scoring in his analysis of variance based on the Q sort (33).

Cluster scoring has an interesting effect on data gathered by means of a Q sort. In this case even though individuals cannot differ in scatter over the total set of items, their subtest profiles can differ widely in scatter. Thus it is possible for some persons to have flat cluster profiles and others to have a high degree of scatter. This results because cluster scores utilize considerably fewer degrees of freedom than are implied in the item profile.

SOME SHORT-CUT FORMULAS

In the course of our investigation, we have discovered the possibility of developing short-cut formulas for studying groups of persons. These

are not entirely satisfactory, because they are based on the average of D^2 over a set of pairs. In general, D provides a better metric than D^2 for studying similarity, since large distances are much magnified in squaring. The following formulas may nonetheless be useful as a first rapid way of answering questions about groups. The formulas also provide insight into the nature of distance measures, since factors which increase mean D^2 will also in general increase mean D and median D. The formulas are particularly useful as a tool for checking computations.

In any group, the mean distance between persons over all pairs of persons in the sample is

$$\overline{D_{ii'}^2} = 2 \frac{N}{N-1} \sum_j V_j. \quad [9]$$

V_j is the variance, equivalent to σ_j^2. This is an expression for the homogeneity of a group or its dispersion. If we take one-half the mean D^2 within the group, we obtain the mean dispersion (distance squared) from the centroid of the sample.

The average D^2 of an individual i from other members of this Group Y, is obtained from

$$\overline{D_{ii'}^2} = \frac{N}{N-1} (O_Y P_i^2 + \sum V_j). \quad [10]$$

Here i' varies over all other persons in Group Y, O_Y is the centroid of the sample, and $O_Y P_i$ is the distance from i to this centroid. O has the coordinates $\bar{x}_j$, the average for j in Group Y. If i is not a member of Y, the coefficient $N/N-1$ is dropped to get the average D^2 from i to all members of Y.

The average D^2 between members of two groups, that is, the average when each member of one group is paired with every member of the other is

$$\overline{D_{ii'}^2} = \overline{O_Y P_{i}^2} + \overline{O_Z P_{i'}^2} + O_Y O_Z^2 (i = 1,$$
$$2, \cdots N_Y; i' = 1, 2, \cdots N_Z). \quad [11]$$

Here we see the average cross-similarity as made up of three components: squared distance between group means, dispersion within the first group, and dispersion within the second group.

CONCLUSIONS

Studies of similarity between sets of scores have used a large number of techniques for assessing similarity. The most satisfactory model appears to be to conceive of the tests as coordinates, and each person's score set as a point in the test space. Then distances between points, computed by the D measure, are an index of similarity between score sets. This measure is a general one, to which other common techniques such as Q correlation can be related. These other techniques frequently disregard or distort some of the information in the data, in ways which may be undesirable in a particular study.

The investigator of similarity must give particular attention to his choice of variates. The similarity measure depends on the content of the variate sets, on the scales used for measuring the variates, on the choice among possible similarity indices, and upon the decision whether to score separate variates or clusters of variates (i.e., subtests). The similarity index gives especially large weight to the first principal component among the scores or items, and therefore may be relatively insensitive to the shape or configuration of profiles. On the other hand, techniques which leave the elevation of the profile out of account are usually undesirable. A formula for a weighted similarity index is offered to reduce any overemphasis on the first component.

Many commonly used operations, including the Q sort and product-moment correlation between persons, ignore differences in scatter between profiles. It is not generally desirable to do this, especially because if any profiles are relatively flat, the similarity indices involving them will be highly unreliable. The loss of information about differences in scatter may also be undesirable on theoretical grounds.

It is most important that any investigator understand the assumptions and limitations of whatever technique he employs to study similarity. Different treatments will yield different conclusions. In many studies, the most appropriate technique will be to apply the formula for D or D_w to profiles based on clusters of items.

Profile research is necessarily faced with many difficulties. In spite of these, it is our hope that the adoption of techniques which include as much information as the data provide, and which do not introduce additional errors of their own, will permit studies of similarity to advance psychological knowledge.

REFERENCES

1. BARNETTE, W. L. Occupational aptitude patterns of selected groups of counseled veterans. *Psychol. Monogr.*, 1951, 65, No. 5 (Whole No. 322).
2. BLOCK, J., LEVINE, L., & McNEMAR, Q. Testing for the existence of psychometric patterns. *J. abnorm. soc. Psychol.*, 1951, 46, 356–359.
3. BURT, C. L. Correlations between persons. *Brit. J. Psychol.*, 1937, 28, 59–96.
4. BURT, C. L. *The factors of the mind.* London: Univer. of London Press, 1940.
5. CALDWELL, BETTY McD., ULETT, G. A., MENSH, I. N., & GRANICK, S. Levels of data in Rorschach interpretation. *J. clin. Psychol.*, 1952, 8, 374–379.

6. CATTELL, R. B. r_p and other coefficients of pattern similarity. *Psychometrika*, 1949, 14, 279–298.

7. CATTELL, R. B. On the disuse and misuse of P, Q, Q_s, and O techniques in clinical psychology. *J. clin. Psychol.*, 1951, 7, 203–214.

8. CATTELL, R. B. The three basic factor-analytic research designs—their inter-relations and derivatives. *Psychol. Bull.*, 1952, 49, 499–520.

9. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1950, 10, 3–31.

10. CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, 16, 297–334.

11. CRONBACH, L. J., & GLESER, GOLDINE. Similarity between persons and related problems of profile analysis. Urbana: Univer. of Illinois, 1952. Tech. Report No. 2, under contract N6ori-07135 with the Bureau of Naval Research (Mimeo.) American Documentation Institute, ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., Document No. 3921, $2.75, microfilm; $7.50, photostats.

12. DuBois, P. H., LOEVINGER, JANE, & GLESER, GOLDINE C. The construction of homogeneous keys for a biographical inventory. Human Resources Research Center, *Research Bulletin*, 1952, 52–18.

13. du MAS, F. M. A quick method for analyzing the similarity of profiles. *J. clin. Psychol.*, 1946, 2, 80–83.

14. du MAS, F. M. On the interpretation of personality profiles. *J. clin. Psychol.*, 1947, 3, 57–65.

15. EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, 16, 407–424.

16. EYSENCK, H. J., Personality. In C. P. Stone (Ed.), *Annual Review of Psychology*. Vol. 3. Stanford: Annual Reviews, 1952. Pp. 151–174.

17. FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, 7, 101–107.

18. FOSBERG, I. A. An experimental study of the reliability of the Rorschach technique. *Rorschach Res. Exch.*, 1941, 5, 72–84.

19. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, 66, No. 18 (Whole No. 350).

20. GAIER, E. L., & LEE, MARILYN C. Pattern analysis: the configural approach to predictive measurement. *Psychol. Bull.*, 1953, 50, 140–148.

21. HARRIS, C. W. Note on profile similarity. Unpublished manuscript.

22. HODGES, J. L., JR. Discriminatory analysis: I. Survey of discriminatory analysis. USAF School of Aviation Medicine, Randolph Field, Texas. 1950.

23. HOLZINGER, K. J. Factoring test scores and implications for the method of averages. *Psychometrika*, 1944, 9, 257–262.

24. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951.

25. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.

26. OSGOOD, C. E., & SUCI, G. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, 49, 251–262.

27. PEARSON, K. On the coefficient of racial likeness. *Biometrika*, 1928, 18, 105–117.

28. RABIN, A. I., & GUERTIN, W. H. Research with the Wechsler-Bellevue test: 1945–1950. *Psychol. Bull.*, 1951, 48, 211–248.

29. RAO, C. R. Tests of significance in multivariate analysis. *Biometrika*, 1948, 35, 58–79.

30. RAO, C. R. The utilization of multiple measurements in problems of biological classification. *J. roy. stat. Soc.*, Sec. B., 1948, 10, 159–203.

31. ROGERS, C. R. The case of Mrs. Oak—a research analysis, *Studies in client-centered psychotherapy*. Psychological Service Center Press, Washington, D.C., 1952, 47–165.

32. STEPHENSON, W. A statistical approach to typology; the study of trait-universes. *J. clin. Psychol.*, 1950, 6, 26–38.

33. STEPHENSON, W. Some observations on Q technique. *Psychol. Bull.*, 1952, 49, 483–498.

34. THOMSON, G. *The factorial analysis of human ability*. (4th Ed.) London: Univer. of London Press, 1950.

35. TYLER, F. T. Some examples of multivariate analysis in educational and psychological research. *Psychometrika*, 1952, 3, 289–296.

36. WARRINGTON, W. G. The efficiency of the Q-sort and other test designs for measuring the similarity between persons. Unpublished doctor's dissertation, Univer. of Illinois, 1952.

37. WEBSTER, H., A note on profile similarity. *Psychol. Bull.*, 1952, 49, 538–539.

HARRISON AND HARRISON'S MODIFICATION OF THE MÜLLER-URBAN WEIGHTS

JOSEPH TIFFIN AND GERALD F. RABIDEAU

Purdue University

In a recent paper Harrison and Harrison (2) suggest that a modification of the Müller-Urban weights used in the psychophysical method of constant stimuli should be generally applicable in psychophysical measurement.

An experiment recently conducted in our laboratories raises the question whether such a modification is worth the additional computational labor. Purposes of the experiment were to examine the respective reliabilities of, and intercorrelation between, the above modification and the classical method. Absolute thresholds for visual acuity (both eyes) were, in the latter case, obtained by using conventional Müller-Urban weights, as described by Guilford (1), and thresholds in the former were computed from the same data using the Harrisons' weighting procedure. In the conventional method, a standard correction for chance successes attributable to guessing was calculated for each stimulus value used.

The visual acuity target utilized in this investigation was the Landolt ring, a nonletter target consisting of a broken ring which may be rotated so as to place the gap at top, bottom, left, or right of the configuration. This target combines minimum separable and minimum visible characteristics as a critical test of acuity. The target sizes are determined by a decimal acuity series, with decimal notation defined as:

$$\text{Decimal acuity notation} = \frac{1}{a},$$

where a equals the angle in minutes subtended by the unit design of the target. In the case of a Landolt ring, unit design is the width of the gap, which is also equal to the width of the stroke. The outside diameter of the ring is five times this distance.

Subject-target distance was 34.5 feet, and surface light on the target was approximately 40 foot-candles.

Thirteen male college students were tested. Each was presented with a series of five consecutive targets, in a descending order of size. The largest target employed was so selected that the subject responded to it correctly less than 100 per cent of the time. Each subject was presented with two series of trials, in each of which the five targets were separately presented 50 times, correct responses (target gap position) being randomly varied. The two series, hereafter referred to as Data A and Data B, were presented in a single session with a separation of a five-minute rest interval.

Subjects were orally instructed to guess when uncertain. Trials were untimed, with the subjects' own rates of response setting the pace throughout the sessions. A change in response was allowed at any time prior to the presentation of the subsequent trial.

The raw data were treated in the two aforementioned manners. In employing the conventional Müller-Urban weights, a correction was made for chance successes in this four-choice test. The formula which was applied was:

$$P_c = P_u - \frac{100 - P_u}{3},$$

where P_c equals corrected percentage correct, and P_u the uncorrected percentage correct. Harrison and Harrison's modification was utilized to fit a second set of curves to the data. It was necessary to compute a table of modified weights for the 25 per cent level of chance success; this was done according to the Harrisons' procedure.

Thresholds obtained are shown in Table 1. The Pearson r for test-retest reliability was .959 for the Urban-Constant method and .943 for the Harrisons' modification. Intercorrelation between the two methods was .995 when thresholds were computed on Data A and B combined.

These data led to the following conclusions:

1. There is a very high correlation ($r = .995$) between the thresholds obtained by the two above methods. Thresholds are, by definition, dependent upon computational methods used to compute them. Therefore there can be as many "true" thresholds as there are computational methods, and one is thus neither more nor less "true" than another.

2. The respective reliabilities of thresholds yielded by these two methods are lacking in any practical differences.

3. Additional computations are necessary to prepare modified Müller-Urban weights according to the Harrisons' procedure. (a) Separate tables of weights must be prepared for each level of chance success encountered in any psychophysical experimentation. (b) Since the Harrison modification results in an asymmetric curve when the Müller-Urban weights are plotted against percentage correct, separate weights must also be computed for *each* percentage, instead

TABLE 1

VISUAL ACUITY THRESHOLDS FOR 13 SUBJECTS OBTAINED
BY TWO COMPUTATIONAL METHODS

Subject Number	Method					
	Urban Constant			Harrisons' Modification		
	Data A	Data B	Data A+B	Data A	Data B	Data A+B
1	1.170	1.206	1.174	1.193	1.180	1.185
2	0.929	0.842	0.910	0.884	0.845	0.864
3	1.007	0.904	0.953	1.010	0.914	0.955
4	1.210	1.289	1.246	1.208	1.279	1.246
5	0.919	0.929	0.921	0.924	0.934	0.926
6	0.926	0.918	0.922	0.938	0.923	0.930
7	1.254	1.230	1.246	1.253	1.231	1.244
8	1.074	1.029	1.052	1.078	1.030	1.054
9	1.396	1.347	1.372	1.401	1.346	1.373
10	1.079	1.043	1.061	1.081	1.045	1.063
11	0.955	1.036	0.992	0.971	1.143	1.038
12	1.517	1.508	1.514	1.539	1.521	1.533
13	0.976	1.005	0.990	0.985	1.008	0.991
Mean	1.109	1.099	1.104	1.113	1.108	1.108

of for only half of the 0 to 100 per cent range, as is the case with the classical weights.

4. When the standard Müller-Urban weights are used in a four-choice situation, the corrected percentage values yield a plot in which the weights are symmetrically distributed, as may be seen in Table G of Guilford's (1) appendix. Percentage correct values range, in this case from 1 to 99.

5. Since both measures yield a measure of precision (h), a standard deviation, and a meaningful estimate of the standard error, it appears that Harrison and Harrison's proposal lacks practicality for wide application.

REFERENCES

1. GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
2. HARRISON, S., & HARRISON, MARGARET J. A psychophysical method employing a modification of the Müller-Urban weights. *Psychol. Bull.*, 1951, 48, 249-256.

BORING, E. G., LANGFELD, H. S., WERNER, H., & YERKES, R. M. (Eds.) *A history of psychology in autobiography.* Vol. IV. Worcester, Massachusetts: Clark Univer. Press, 1952. Pp. viii+356. $7.50.

The first three volumes in this series were published in 1930, 1932, and 1936. As a result of action taken by the American Psychological Association we now have Vol. IV, which includes the autobiographies of Walter Van Dyke Bingham, Edwin Garrigues Boring, Cyril Burt, Richard M. Elliott, Agostini Gemelli, Arnold Gesell, Clark L. Hull, Walter S. Hunter, David Katz, Albert Michotte, Jean Piaget, Henri Piéron, Godfrey Thomson, L. L. Thurstone, and Edward C. Tolman. The ratio of Americans to Europeans is 8 to 7, as compared with 7 to 8 in Vol. I. At the time the authors were selected (13 of them early in 1950, the other two a year later), all but Piaget were above the age of 60 and nearing the end of their professional careers. Three have since died, Hull and Bingham before the volume was published and Katz shortly thereafter. The relative maturity of the authors is one of several reasons why Vol. IV is the most interesting and valuable of the series. Some of the surviving authors of Vols. II and III who were caught in mid-career will regret that they were not saved for a later volume. Too early selection may leave half or a third of the life story untold; and what is worse, it leaves the author tagged forever with views he may have later discarded. This reviewer shudders to recall some of the dogmatic opinions he expressed in Vol. II, particularly one concerning racial differences in intelligence. Burt had the good sense to refuse an invitation he received at about the same time.

There will be disagreement on the Committee's selection of authors, for no such list could be expected to meet with universal approval. There will probably be little disagreement in this country on the European group, except possibly in the case of Gemelli, whose work is little known on this side of the Atlantic. However, psychologists in both hemispheres will wonder about the absence of Köhler and Lashley in the American list. Their omission, the reviewer is reliably informed, was not the fault of the Committee; both were invited to contribute and both firmly refused. But other omissions will also be questioned, among them the following past presidents of the APA, all of whom were above the age of 60 and still active professionally: Dashiell, Guthrie, Langfeld, and Miles (to name them alphabetically). These, as well as several others who might have been considered, were passed over in favor of Elliott, who has done almost no research, and Bingham, who is less known for his research than for other activities. Yet both of these choices can be defended. Elliott in his long career as a brilliant teacher and departmental administrator has helped mold the careers of many able psychologists; Bingham, besides making important contributions to military psychology in two world wars, probably influenced the early history of industrial psychology in this country more than anyone else.

It is impossible to summarize here the individual autobiographies, and

477

to attempt an appraisal of their relative merits would be presumptuous. The reviewer has chosen instead to comment on the unusual aspects of some of the careers and to point out differences among authors in the ways they approached their task and in the kinds of information they have given or omitted.

The volume is notable for the number of authors who were migrants to psychology from other fields. Of course such migration is not a recent phenomenon; most of the founders of modern psychology were migrants, usually from philosophy but sometimes from other disciplines. About half of our 15 authors got into psychology in roundabout ways. Boring, Thurstone, and Tolman were engineering graduates, Thomson took his doctorate in physics, Gemelli in both medicine and biology, and Piaget in the natural sciences. Hull, too, entered psychology as a migrant but at an earlier stage in his education. The total impact of these men on contemporary psychology has surely been very great, but whether less or more than it would have been if they had entered psychology by a more direct route is impossible to say. On the one hand, some of them were able to apply their previous training in science to certain types of psychological problems more effectively than if they had not detoured; on the other hand, the detours in some cases were costly in time. Here are a few facts about the detours and delays experienced by this group.

Boring had only one course in psychology (that fortunately with Titchener) prior to age 24 when, after returning to Cornell for graduate work in physics, he was "caught" for a course in animal psychology that sealed his future. His Ph.D. at 28 does not suggest much retardation, but the fact is that he had spent four years as a student of engineering and another year working in a steel mill.

Gemelli, who took his M.D. degree at the age of 24 and his Ph.D. in biology at 26, was 31 when he first became interested in experimental psychology and close to 40 when this interest became paramount.

Tolman lost two years from illness and was 25 years old when he graduated in engineering at Massachusetts Institute of Technology. He had gone to M.I.T. not because he wanted to be an engineer but because of family pressure; his father had graduated from that institution and was one of its trustees. During his senior year he read some of William James and fancied he wanted to become a philosopher. He entered Harvard that summer for a course in philosophy and one in animal behavior. The latter course turned him to psychology, in which he got his doctorate four years later at the age of 29.

Thomson received his Ph.D. in physics and mathematics at age 25 and first became interested in psychology as a result of having to give a course in educational psychology to prospective teachers. When he was 30 years old he visited Myers' laboratory, where he was allowed to spend a long summer vacation working by himself through the experiments in Myers' textbook. Returning to his teaching he undertook some research on the side and before the age of 33 had his first two papers ready for publication, one on psychophysics, the other on factor analysis. The latter created a sensation when it was published two years later under the title "A Hierarchy without a General Factor." Thomson's case is unique in that he never had a real course in psychology, though he was awarded a D.Sc. in the subject by

Cambridge University on presentation of his researches and on passing an oral examination. His age at the time was 32.

Piaget took his doctorate in science at the age of 22 with a dissertation on mollusks. By that time he had decided upon psychology and for the next three years studied it first at Zurich then at the Sorbonne. Though a migrant to psychology he was not delayed, for he got his degree earlier than is usual and has managed to utilize in his psychology some of the biological concepts he had developed in his late teens.

Hull entered college at the age of 22, after two years of teaching in a rural school, and devoted most of his freshman and sophomore years to courses in mathematics, physics, and chemistry in preparation for a major in mining engineering. Then he was stricken with polio, which cost him an entire year and left him so badly crippled that he had to look about for a more sedentary occupation. While trying to reach a decision he taught in a village school for two years, read James' *Principles*, and at age 28 began his junior undergraduate year in college as a psychology major. He was 31 when he began his graduate work and 35 when he received his doctorate.

The reviewer finds the autobiographies of these migrants to psychology fascinating and provocative of questions not only on the merits of vocational guidance but also on the kinds of training useful in the making of psychologists. Readers will be amused by (and some may even profit from) Thurstone's caustic comments on the differences between his professors of engineering and some of his professors of psychology. One is led to wonder whether undergraduate courses in the physical or biological sciences may not provide a better foundation for the psychologist-to-be than do the usual undergraduate courses in psychology. At that level, perhaps, the acquisition of psychological information may be less important than training in the logic and techniques of the more exact sciences.

Gesell was not a migrant to psychology though he did make some detours before settling down to his research career. First, he taught in high schools for three years prior to taking his doctorate in psychology at age 26; then, after four additional years of teaching (three of them in normal schools), he decided that the lifework he had planned called for a medical degree, and he spent five years getting it. Thus he devoted twelve years to teaching and study over and above the time required for the Ph.D. degree. Perhaps not even Gesell could say how many (if any) of these years were really time lost; if loss there was, it seems to have been made good later by an extraordinary tempo of research productivity.

All the autobiographies have certain things in common: each tells something about the immediate circumstances that influenced the author to become a psychologist, traces the leading developments in his work, and indicates what he thinks his more important contributions have been. Apart from this common core, which varies from author to author in amount of detail and in distribution of emphasis, the autobiographies present striking contrasts. Some are almost completely intellectualistic, others are intimate and personal. Michotte illustrates the former tendency; Boring, Burt, and Tolman the latter. A majority give some information, and several of them a great deal, about ancestry, family back-

design considerations of the aeronautical engineer. The prediction that it would soon become a classic in the field has been realized. McFarland has now turned his attention to how human factors influence the efficient operation of air transportation. The new book is a necessary addition to the complete aeromedical bookshelf and may confidently be expected to achieve the success of its companion volume.

The introduction states that: "The primary objective of this book is to improve the safety, efficiency, economy, and comfort of airline operations. It is the author's conviction that air transports can be operated efficiently with minimum risk only in so far as the human variables are understood and controlled." It is also McFarland's conviction that, "Greatest improvement can be achieved ... by strengthening the medical, safety and personnel programs of the airlines."

It should be pointed out at the outset that this work was not written primarily for psychologists. Those who expect a treatise on aviation psychology, particularly the human engineers, are likely to be disappointed. The industrial psychologists will fare better since there is much that is of general interest to the applied psychologist, and five or six chapters on selection, training, fatigue, aging, etc. are of especial interest. The largest portion of the book is devoted to aviation medicine, more exactly *preventive* medicine in airline operations, and it is addressed rather pointedly to the air transport industry. However, the book is of far broader scope, as the author makes clear in his preface and choice of subtitle. Much of the material on industrial hygiene is applicable to occupational health and safety in other industrial fields.

The book contains an enormous amount of information and is thoroughly documented with over 1200 references. This material is drawn from the most diverse sources and has been painstakingly analyzed for its practical significance to airline operations. We have come to expect from the author the ability to put selections from such a mass of material into a form suitable for practical use. Indeed, as was true of his first book, an outstanding feature of McFarland's presentation is that it contains so many specific and down-to-earth recommendations.

The main theme of the book is safety, and the author starts with *preventive* safety in the air and on the ground. His first concern is with the pilots and flight crews in whose hands the safety of the public in part rests. He treats their psychological and medical selection, their training and indoctrination, the maintenance of their health, the necessity of periodic medical examinations, the control of fatigue, and the factors in their lives that affect efficiency and aging. Next he undertakes a similar analysis of the selection, placement, health, and safety of the supporting people on the ground. Since there are 10 people on the ground for every pilot in the air, the sickness and accident rates of ground personnel are of great economic importance in airline operations. Ground conditions also bear upon safety in the air, since the two are intimately related. Sanitary control and the prevention of infectious diseases are especially important to operations in certain overseas bases, and careful attention is given to this problem as it affects operating personnel, passengers, the transportation of diseases, and quarantine procedures. Finally, an analysis is made of the factors that operate in keeping passengers well and happy. This includes a discussion of the

cial problems that arise in the air nsportation of patients.

A second aspect of safety relates *rescue* safety, and the author has de an excellent study of aircraft cidents resulting in crashes or ching. Search and rescue practices, d survival and emergency proce- res are treated in full detail.

Throughout each chapter the im- rtance of health and safety meas- es in the management of airlines is peatedly emphasized. The book ncludes, therefore, with an analysis health and safety programs that ould be developed not only in the terest of the flying public but in the terest of the airlines. McFarland akes a good case for his point that ell-organized medical services are ofitable as well as desirable.

Occasionally the material in this ook touches upon the same ground s the previous book. Where this is ue, new material and a new em- hasis have been added. The style is ood and the material is set in large ype on large, easy-to-read, double- olumn pages. With an eye to use as handbook, as well as for clarification the text, there are 171 tables, 156 gures, and an appendix of terms equently used in the aircraft in- ustry. Each chapter concludes with full summary and recommendations r improvement in airline operations

J. W. GEBHARD.

The Johns Hopkins University.

UNDIN, ROBERT W. *An objective psychology of music.* New York: Ronald, 1953. Pp. ix + 303. $4.50.

Up to the present the more impor- ant psychological texts dealing with musical problems have been Sea- hore's *Psychology of Music* (McGraw- Iill, 1938), Diserens and Fine's *A sychology of Music* (Authors, 1937), Mursell's *The Psychology of Music* (Norton, 1937), and Schoen's *The*

Psychology of Music (Ronald, 1940). Max Meyer's *The Musician's Arith- metic* (Ditson, 1929), though written as a text, has had a deplorably limited audience. It contains many exciting neurological speculations but is extremely difficult to read.

Seashore's book covers little more than the work of his own laboratory. It is distinctly hereditarian in flavor and favors the natural science ap- proach. Seashore later attempted to broaden his coverage somewhat with his *Why We Love Music* (Ditson, 1941) and his *In Search of Beauty in Music* (Ronald, 1947).

The Diserens and Fine text has a better coverage but devotes what many consider a disproportionate amount of space to topics such as the possible origin of music, animal au- ditors, mythology and folklore, and the relation of music to magic and sorcery and to religion, melancholy, and ecstasy. The Mursell offering has been perhaps the most palatable to musicians, though psychologists have at times been distressed at the extreme Gestalt interpretation given all phenomena. Schoen's well-round- ed text has the hereditarian bias of his teacher, Seashore.

Lundin's addition to this shelf of books considers most of the material treated in the earlier volumes and brings all topics reasonably up to date. A surprising omission, how- ever, concerns Pratt's *The Meaning of Music* (McGraw-Hill, 1931). Other psychological books seemingly ignored are Schoen's *The Under- standing of Music* (Harper, 1945) and the above-mentioned texts by Diserens and Fine and by Meyer.

Lundin looks to cultural explana- tions rather than to nativistic ones. Thus, consonance is regarded not as a phenomenon of the natural proper- ties of stimuli but as the product of a particular culture. Music tests are

not thought to measure innate capacities. Because of their present low level of efficiency Lundin takes a dim view of them although he himself has developed a few.

Noticeable particularly in the first chapter is the interbehavioral stand of J. R. Kantor, to whom the book is dedicated. This theoretical orientation will probably confuse only slightly the psychologist who is unacquainted with Kantonian theory.

Lundin has done a workman-like job throughout the volume's sixteen chapters. He ends with a most usable bibliography of over 260 titles. In the reviewer's opinion this book should supersede the earlier texts and should greatly aid the occasional teacher who ventures to give a course in the psychology of music.

PAUL R. FARNSWORTH.
Stanford University.

KUHLEN, RAYMOND G., & THOMPSON, GEORGE G. (Eds.) *Psychological studies of human development.* New York: Appleton-Century-Crofts, 1952. Pp. xiv+533. $3.50.

This is a well-organized collection of 71 studies that have appeared in 11 books and 33 journals. More than a third of the articles appeared in a group consisting of the *Journal Educational Psychology* and the *Journal of Abnormal and Social Psychology* (6 articles), *Child Development* and the *Journal of Genetic Psychology* (5 articles), and *Genetic Psychology Monographs* (4 articles). All articles were published in America, and all within the past 25 years.

The reports of 104 investigators are divided into 14 groupings: physical factors, learning and adjustment, sociocultural conditions, intellectual changes with age, intelligence, language-conceptual growth, interest patterns, social values and attitudes, interpersonal relationships, home and family relations, vocational orientations, and factors in personal and emotional adjustment.

In the opinion of this reviewer, the emphasis on educational and experimental studies gives this symposium a cast that may bolster its standing as a reference text for courses in child psychology but also inhibit its appeal for people hoping to achieve practical goals with children and their parents.

T. W. RICHARDS.
Louisiana State University.

BARKER, R. G., WRIGHT, B. A., MEYERSON, L., & GONICK, M. R. *Adjustment to physical handicap and illness: a survey of the social psychology of physique and disability.* New York: Soc. Sci. Res. Coun., 1953. Pp. vi+440. $2.00.

BERKMAN, TESSIE D. *Practice of social workers in psychiatric hospitals and clinics.* New York: American Association of Psychiatric Social Workers, 1953. Pp. ix+158. $2.00.

BUROS, OSCAR KRISEN. (Ed.) *The fourth mental measurements yearbook.* Highland Park, N.J.: Gryphon, 1953. Pp. xxiv+1163. $18.00.

Conferences on drug addiction among adolescents. Sponsored by Committee on Public Health Relations of the New York Academy of Medicine. New York: Blakiston, 1953. Pp. xvi+320. $4.00.

CRAIG, ROBERT C. *The transfer value of guided learning.* New York: Bureau of Publications, Teachers College, Columbia Univer., 1953. Pp. viii+85. $2.75.

CROW, LESTER D., & CROW, ALICE. *Child psychology.* New York: Barnes & Noble, 1953. Pp. xv+267. $1.50.

EDUCATIONAL RECORDS BUREAU. *1953 achievement testing program in independent schools and supplementary studies.* Bulletin No. 61. New York: Educational Records Bureau, 1953. Pp. xii+86. $1.50.

EVANS, RICHARD I. *A telecourse guide to introductory psychology.* Houston, Texas: Univer. of Houston Press, 1953. Pp. vii+102.

GOLDHAMER, HERBERT, & MARSHALL, ANDREW W. *Psychosis and civilization.* Glencoe, Ill.: Free Press, 1953. Pp. 126. $4.00.

HALPERN, FLORENCE. *A clinical approach to children's Rorschachs.*

New York: Grune & Stratton, 1953. Pp. xiii+270. $6.00.

HARARY, FRANK, & NORMAN, ROBERT Z. *Graph theory as a mathematical model in social science.* Ann Arbor: Univer. of Michigan Press, 1953. Pp. v+45. $1.00.

HATHAWAY, STARKE R., & MONACHESI, ELIO D. (Eds.) *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. of Minnesota Press, 1953. Pp. vi+153. $3.50.

HEINICKE, CHRISTOPH. *Bibliography on personality and social development of the child.* WHITING, BEATRICE. *Selected ethnographic sources on child training.* New York: Soc. Sci. Res. Coun., 1953. Pp. vii+130. $1.00.

HULETT, J. E., Jr., & STAGNER, ROSS. (Eds.) *Problems in social psychology: an interdisciplinary inquiry.* Urbana: Univer. of Illinois Press, 1952. Pp. viii+271. $2.50.

JONES, MAXWELL. *The therapeutic community.* New York: Basic Books, 1953. Pp. xxi+186. $3.50.

JUNG, C. G. *Psychology and alchemy* (Vol. 12 of collected works). New York: Pantheon, 1953. Pp. xxiii+563. $5.00.

KARPF, FAY B. *The psychology and psychotherapy of Otto Rank.* New York: Philosophical Library, 1953. Pp. ix+129. $3.00.

KNAPP, ROBERT H. *Practical guidance methods: for counselors, teachers and administrators.* New York: McGraw-Hill, 1953. Pp. xi+320. $4.25.

KOGAN, L. S., HUNT, J. McV., & BARTELME, PHYLLIS F. *A follow-up study of the results of social case work.* New York: Family Service Association of America, 1953. Pp. 115. $2.50.

LACEY, OLIVER L. *Statistical meth-*

ods in experimentation. New York: Macmillan, 1953. Pp. xi+249. $4.50.

LINDGREN, HENRY CLAY. *Psychology of personal and social adjustment.* New York: American Book Co., 1953. Pp. ix+481. $4.50.

MORENO, J. L. *Who shall survive?* New York: Beacon, 1953. Pp. cxiv+ 763. $10.00.

MOWRER, O. HOBART. *Psychotherapy: theory and research.* New York: Ronald, 1953. Pp. xviii+700. $10.00.

MURPHY, GARDNER. *In the minds of men: the study of social tensions in India.* New York: Basic Books, 1953. Pp. xiv+306. $4.50.

NUTTIN, JOSEPH. *Psychoanalysis and personality; a dynamic theory of normal personality.* New York: Sheed and Ward, 1953. Pp. xiv+310. $4.00.

PATTY, WILLIAM L., & JOHNSON, LOUISE S. *Personality and adjustment.* New York: McGraw-Hill, 1953. Pp. viii+403. $4.75.

PATTY, WILLIAM L., & JOHNSON, LOUISE S. *Instructor's manual (to accompany Personality and adjustment).* New York: McGraw-Hill, 1953. Pp. 64. Free.

SULLIVAN, HARRY S. *The interpersonal theory of psychiatry.* (Helen S. Perry and Mary L. Gawel, Eds.) New York: Norton, 1953. Pp. xviii+ 393. $5.00.

PETERS, R. S. (Ed.) *Brett's history of psychology.* New York: Macmillan, 1953. (Abridgement of earlier 1912–1921 3-vol. work.) Pp. 742. $7.50.

PHILLIPS, LESLIE, & SMITH, JOSEPH G. *Rorschach interpretations: advanced technique.* New York: Grune & Stratton, 1953. Pp. xi+385. $8.75.

PODOLSKY, EDWARD. *Encyclopaedia of aberrations.* New York: Philosophical Library, 1953. Pp. viii+ 550. $10.00.

RECKTENWALD, LESTER N. *Guidance and counseling.* Washington, D. C.: Catholic Univer. of America Press, 1953. Pp. xiv+192. $3.2 (cloth ed.), $2.50 (paper cover)

ROE, ANNE. *The making of a scientist.* New York: Dodd, Mead & Co, 1952, 1953. Pp. ix+244. $3.75.

SHERIF, MUZAFER, & SHERIF, CAROLYN W. *Groups in harmony and tension.* New York: Harper, 1953. Pp. xiii+316. $3.50.

STEINER, LEE R. *A practical guide for troubled people.* New York: Greenberg, 1952. Pp. 299. $3.50.

TAKALA, MARTTI. *Studies of psychomotor personality tests I.* Helsinki: Finnish Academy of Science and Letters, 1953. Pp. 130. 400 mk.

TOWNSEND, JOHN C. *Introduction to experimental method.* New York: McGraw-Hill, 1953. Pp. ix+220. $4.00.

VERNON, PHILIP E. *Personality tests and assessments.* London: Methuen, 1953. Pp. xi+220. 18 s.

VON FOERSTER, HEINZ. (Ed.) *Cybernetics: circular causal and feedback mechanisms in biological and social systems.* (Transactions of the ninth conference). New York: Josiah Macy, Jr. Foundation, 1953. Pp. xx+184. $4.00.

WATSON, DAVID L. *The study of human nature.* Yellow Springs, Ohio: Antioch Press, 1953. Pp. x+262. $3.50.

WHITING, JOHN W. M., & CHILD, IRVIN L. *Child training and personality: a cross-cultural study.* New Haven: Yale Univer. Press, 1953. Pp. vi+353. $5.00.

WILKIE, J. S. *The science of mind and brain.* New York: Longmans, Green, 1953. Pp. viii+160. Text ed. $1.80. Trade ed. $2.25.

WITTKOWER, ERIC, & RUSSELL, BRIAN. *Emotional factors in skin disease.* New York: Harper, 1953. Pp. x+214. $4.00.

INDEX OF AUTHORS

GENERAL REVIEWS, SHORT ARTICLES, SPECIAL REVIEWS, NOTES

BOOKS REVIEWED

BOOK REVIEWERS